

(12)特許協力条約に基づいて公開された国際出願

(19)世界知的所有権機関
国際事務局



(43)国際公開日
2003年1月23日 (23.01.2003)

PCT

(10)国際公開番号
WO 03/007187 A1

(51)国際特許分類⁷: G06F 17/30, 17/50, G01N
33/566, A61K 38/00, 45/00, C12N 15/00

LTD.) [JP/JP]; 〒541-0045 大阪府 大阪市 中央区道修町四丁目 1番 1号 Osaka (JP).

(21)国際出願番号: PCT/JP02/07057

(72)発明者; および

(22)国際出願日: 2002年7月11日 (11.07.2002)

(75)発明者/出願人(米国についてのみ): 井ノ岡博 (INOOKA,Hiroshi) [JP/JP]; 〒631-0022 奈良県 奈良市 鶴舞西町2丁目10番E-203号 Nara (JP). 山本善雄 (YAMAMOTO,Yoshio) [JP/JP]; 〒666-0261 兵庫県 川辺郡 猪名川町松尾台2丁目1番地2 D-103 Hyogo (JP).

(25)国際出願の言語: 日本語

(74)代理人: 小林浩, 外 (KOBAYASHI,Hiroshi et al.); 〒104-0028 東京都 中央区 八重洲2丁目8番7号 福岡ビル9階 Tokyo (JP).

(26)国際公開の言語: 日本語

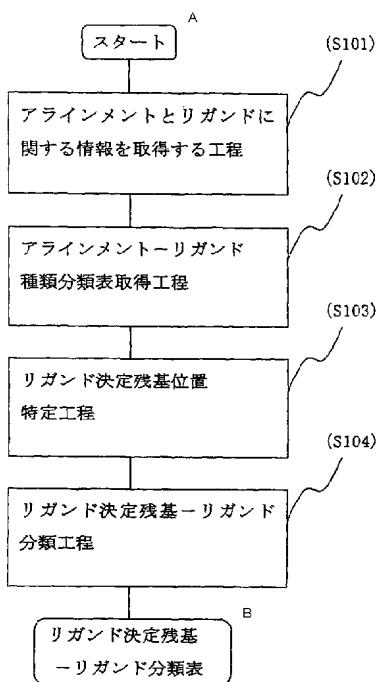
(30)優先権データ:
特願2001-212749 2001年7月12日 (12.07.2001) JP

(71)出願人(米国を除く全ての指定国について): 武田薬品工業株式会社 (TAKEDA CHEMICAL INDUSTRIES,

[続葉有]

(54)Title: METHOD OF PRESUMING LIGAND AND METHOD OF USING THE SAME

(54)発明の名称: 結合分子予測方法およびその利用方法



(57)Abstract: A method for estimating a ligand or a ligand type directly binding (or coupling) to GPCR based on GPCR sequential data and, in its turn, estimating its function. More specifically, a method of presuming a ligand of a protein with unknown ligand which comprises obtaining classification data of proteins with known ligands wherein the alignments of the proteins with known ligands correspond respectively to the ligands or ligand types, obtaining ligand-determining residue-ligand classification data which shows the correlations among ligand-determining residues and ligands or ligand types to thereby obtain the alignments of proteins with unknown ligands, applying the data at least concerning the ligand-determining residues in the alignments of the proteins with unknown ligands as described above to the ligand-determining residue-ligand classification data, and thus estimating the ligand or ligand type of the protein with the unknown ligand, etc.

A ...START
(S101)...STEP OF OBTAINING DATA CONCERNING ALIGNMENTS AND LIGANDS
(S102)...STEP OF OBTAINING ALIGNMENT-LIGAND TYPE CLASSIFICATION LIST
(S103)...STEP OF SPECIFYING LIGAND-DETERMINING RESIDUE POSITIONS
(S104)...STEP OF CLASSIFYING LIGAND-DETERMINING RESIDUE-LIGAND B...LIGAND-DETERMINING RESIDUE-LIGAND CLASSIFICATION LIST

[続葉有]



(81) 指定国(国内): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

特許(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI 特許(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:
— 國際調査報告書

(84) 指定国(広域): ARIPO 特許(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア特許(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイドスノート」を参照。

(57) 要約:

本発明は、G P C R の配列情報から直接的にG P C R に結合する(、あるいは共役する)分子又は分子の種類を予測し、その機能を予測することにつながる方法を提供する。具体的には、本発明は、結合分子既知タンパク質のアラインメントと、結合分子又は結合分子の種類とを対応付けた結合分子既知タンパク質分類情報を得た後、結合分子決定残基と結合分子又は結合分子の種類との相関関係を表す結合分子決定残基-結合分子分類情報を得て、結合分子未知タンパク質についてのアラインメントを得る工程と、前記結合分子未知タンパク質のアラインメントのうち少なくとも結合分子決定残基についての情報を、結合分子決定残基-結合分子分類情報に当てはめ、結合分子未知タンパク質の結合分子又は結合分子の種類を予測する結合分子未知タンパク質の結合分子予測方法、などを提供する。

明細書

結合分子予測方法およびその利用方法

技術分野

5 本発明は、結合分子未知タンパク質の結合分子予測方法、該予測方法を用いた医薬の製造方法、及び結合分子未知タンパク質の結合分子を予測するためのコンピュータに関し、より詳しくは、結合分子既知タンパク質のアミノ酸配列（又はシークエンスアラインメント）とその結合分子又は結合分子の種類とに関する情報から結合分子未知タンパク質の結合分子又は結合分子の種類を予測する方法、
10 該方法を用いた医薬、及び当該方法に用いられるコンピュータに関する。さらに本発明は、結合分子や結合分子の種類を予測することを通して結合分子が未知であるタンパク質の機能を予測する方法等に関する。

背景技術

15 ヒトゲノムの配列情報が公開され（Nature 第409巻、第6822号（2001）；Science 第291巻、第5507号（2001））、既知の遺伝子も含め3-4万の遺伝子の存在が報告された。しかし、遺伝子と細胞内機能の関係が未知のものが数多く存在している。特に、疾病・疾患の原因となる遺伝子に関する情報はその病態を明らかにする上で重要であるばかりでなく、そ
20 の診断・予防・治療に用いる医薬を開発する場合にも必要不可欠なものである。

医薬開発の標的となりうる分子は、その分子を含む医薬が体内に入ったとき直接または間接的に他の分子と結合するような分子であり、ほとんどの場合がタンパク質と考えられる。このタンパク質としては低分子物質を結合するタンパク質、低分子または高分子の基質に対して触媒活性を有する酵素、高分子物質を結合するタンパク質などが挙げられる。その中でも低分子物質を結合するタンパク質を標的とする医薬品が非常に多く存在する。標的分子として特に注目されるタンパク質として、膜タンパク質としてシグナルの伝達に関与するGタンパク質共役型受容体タンパク質があげられる。

膜タンパク質の例では、細胞外から細胞内へのシグナルの伝達は、膜貫通ドメ

インを有する膜タンパク質がその膜貫通ドメインで構造変化を生ずることにより進行する。膜貫通ドメインを有する上記のGタンパク質共役型受容体タンパク質(G protein coupled receptor protein, 以下G P C Rと略すことがある)は7回膜貫通ドメインをその分子内に有するシグナル伝達に関与する膜タンパク質である。G P C Rは細胞外側にN末端を有し、7回膜貫通ドメインを経て細胞内にC末端を有するトポロジーを持ち、リガンドはその種類によってN末端領域や膜貫通領域に結合する。また、細胞内のC末端部分と細胞内ループ2にGタンパク質が結合することが知られている。現在までに唯一G P C Rファミリーの一種であるロドプシンの結晶構造解析が行われている(*Palczewski et al. Science* 第289巻、739-745頁(2000))が、その他のG P C Rの結晶構造解析に関する報告はない。また各種生物学実験により、以下のようにG P C Rの活性化が起きることが明らかとなってきた。即ち、(i) 受容体にリガンドが結合し、G P C Rの構造の変化する。(ii) 共役Gタンパク質にG P C Rの構造変化が伝達され、Gタンパク質の一部が遊離する。(iii) 情報(シグナル)が細胞内へ伝達される。

受容体と特異的に結合する分子はリガンドと呼ばれる。G P C Rの場合、既知のリガンド分子としては、ドーパミン、セロトニン、メラトニン、ヒスタミンのような生体アミン、プロスタグランジン、ロイコトリエンなどの脂質誘導体、核酸、グルタミンなどのアミノ酸、アンギオテンシン、セクレチン、ソマトスタチンのような生理活性ペプチド類などがあげられる。

Gタンパク質はシグナル伝達系においてトランスデューサーとして機能し、 α 、 β 、 γ の3種のサブユニットから構成される。これらのうち α サブユニットはG T Pへの結合およびG T Pを加水分解する活性を有し、各々のGタンパク質に特有のサブユニットである。即ち、共役するG α タンパク質の種類により、G P C Rの機能が決定されることになる。現在までに、G α タンパク質としてはG_s、G_i、G_oおよびG_tが単離精製され、その性質が調べられている。

Gタンパク質共役型レセプタータンパク質は生体の細胞や臓器の各機能細胞表面に存在し、それら細胞や臓器の機能を調節する分子、例えば、ホルモン、神経伝達物質および生理活性物質等の標的として生理的に重要な役割を担っている。

レセプターは生理活性物質との結合を介してシグナルを細胞内に伝達し、このシグナルにより細胞の賦活や抑制といった種々の反応が惹起される。

各種細胞や臓器における複雑な機能を調節する物質と、その特異的レセプタタンパク質、特にGタンパク質共役型レセプタタンパク質との関係を明らかにすることは、各種細胞や臓器における複雑な機能を解明し、それら機能と密接に関連した医薬品開発に非常に重要な手段を提供することとなる。

近年、生体内で発現している遺伝子を解析する手段として、cDNAの配列をランダムに解析する研究が活発に行なわれており、その結果として得られたcDNAの断片配列がExpressed Sequence Tag (EST) としてデータベースに登録され、公開されている。しかし、多くのESTは配列情報だけを提供し、配列が有する機能を推定することは困難である。

Gタンパク質共役型レセプターはその全てが見出されているわけではなく、現時点でもなお、未知のGタンパク質共役型レセプター、また対応するリガンドが同定されていない、いわゆるオーファンレセプターが多数存在しており、新たなGタンパク質共役型レセプターの探索および機能解明が切望されている。

ゲノムの配列情報が公開されたことにより、既知の配列情報と比較することによって何らかの機能を有すると推測できる遺伝子領域を調べることが可能となつた。遺伝子の塩基配列間もしくはそれが翻訳されたタンパク質のアミノ酸配列間の類似性に基づき、新規タンパク質の結合分子や機能の予測が実施されている。

上記のように、配列の類似性を計算し、配列間の適切な対応関係(シークエンスアラインメント)を表示するソフトとしてClustal WやBLASTなどが用いられている。これらのソフトから得られたシークエンスアラインメントを用い、類似性解析によって、新規タンパク質にはその類似タンパク質と類似の機能があるという概念に基づき、機能予測が実施されている。このような類似性に基づく予測は、基本的な機能の予測には貢献するが、基本的な機能が同じタンパク質グループに対する結合分子の予測に関しては効果が乏しい。例えば、その新規タンパク質がGPCRかどうかの分類化には役立つが、類似度が極めて高いものでない限り、そのGPCRに結合するリガンドや共役Gタンパク質の種類の予測には向きである。実際にGPCRに結合する未知のリガンド分子を検索・決定するためには、細胞

抽出物などを利用して多くの候補化合物の活性測定を行う必要があるが、物性の異なる結合分子の候補化合物は同時に評価することができないという問題が生じる。そこで、複雑な活性測定などを行うことなく、リガンド分子の種類を予測することができたとすれば、効率的なリガンド分子の決定に結びつくこととなる。

したがって、タンパク質の配列情報を類似性とは異なる概念で捉えることが必要となる。類似性によって分類されたタンパク質グループ(例えばGPCR)内の各メンバーに対して、各タンパク質に結合する分子を予測することができれば、リガンドの同定やそのタンパク質とリガンドの機能解析に有用である。のために、タンパク質の1次元的な配列情報から立体構造を予測し、計算機上での結合実験によって結合分子を予測する検討も実施されている。しかし、現状ではタンパク質の立体構造の予測および計算機上での結合実験の精度がないために、実用的でない。これらの精度の低さと煩雑さを回避し、配列情報から直接結合分子や機能について検討できることは本発明の技術分野において有用なことである。

配列情報から、結合分子や機能を予測する方法としては、例えば、限定された既知のGPCRに対して、共役するG α タンパク質の種類の決定に関与すると思われるアミノ酸の変異(ならびにその変異位置)についての報告がなされた(Bulseco and Schimerlik, Molecular Pharmacology, 49: 132-141 (1996); Burstein, Spalding and Brann, Biochemistry, 37: 4052-4058 (1998); Kazmi et al., Biochemistry, 39: 3734-3744 (2000))が、新規GPCRに対して共役するG α タンパク質の種類を予測するような報告は知られていない。

また、やはりGPCRに関して、GPCRとリガンド双方のアミノ酸の変異を解析することにより、リガンドとそのGPCRの結合に関与する部位を予測する方法、即ちcorrelated mutation analysis(CMA)法が開発された(Singer et al., Receptors and channels, 3: 89-95 (1995))が、GPCRの配列を単独で用いて未知のGPCRの結合分子(及び/又は結合分子の種類)を選択した報告例はない。

以上のように、例えばGPCRの場合、GPCRの配列情報から直接的にGPCRに結合する(、あるいは共役する)分子又は分子の種類を予測し、その機能を予測することにつながる方法は知られていない。

また、やはり GPCR についていえば、新たな GPCR を発見した場合、その GPCR を利用した医薬を製造するためには、当該 GPCR の機能を把握した上で、当該 GPCR に結合する（あるいは共役する）分子又は分子の種類を実験等により把握しなければならず、莫大な費用と手間がかかるという問題がある。

5

発明の開示

本発明は、機能未知の配列情報に対し、それに結合できる分子等を簡便かつ精度よく予測する方法、当該方法を用いた医薬の製造方法及び、これらに用いられるコンピュータに関する。

10 上記課題の少なくとも一つは、以下の発明によって解決される。

(1) 結合分子未知タンパク質に結合する結合分子を予測する結合分子未知タンパク質の結合分子予測方法であって、アミノ酸配列と結合分子とが既知である結合分子既知タンパク質について、少なくとも 2 以上の結合分子既知タンパク質のシークエンスアラインメントと、結合分子又は結合分子の種類とを対応付けた結合分子既知タンパク質分類情報を得る工程と、前記結合分子既知タンパク質分類情報を用いて、結合分子既知タンパク質のシークエンスアラインメントの位置のうち結合分子を決定することに関与すると想定される位置である結合分子決定残基位置を 1 又は 2 以上特定する工程と、前記結合分子決定残基位置におけるアミノ酸残基（結合分子決定残基）と、結合分子又は結合分子の種類とを対応付けることにより、結合分子決定残基と結合分子又は結合分子の種類との相関関係を表す結合分子決定残基—結合分子分類情報を得る工程と、前記結合分子既知タンパク質と同じ種類の結合分子未知タンパク質について前記結合分子既知タンパク質間のシークエンスアラインメントに対して結合分子未知タンパク質の配列を整列させ、結合分子未知タンパク質のシークエンスアラインメントを得る工程と、前記結合分子未知タンパク質のシークエンスアラインメントのうち少なくとも結合分子決定残基についての情報を、結合分子決定残基—結合分子分類情報に当てはめ、結合分子未知タンパク質の結合分子又は結合分子の種類を予測する工程とを含む結合分子未知タンパク質の結合分子予測方法。

(2) 前記結合分子が、リガンド、調節因子、エフェクター、補酵素のいずれ

かである上記（1）に記載の結合分子未知タンパク質の結合分子予測方法。

（3）前記結合分子が、2以上の種類に分類され、当該分類された結合分子の種類を予測する上記（1）又は（2）に記載の結合分子未知タンパク質の結合分子予測方法。

5 （4）結合分子未知タンパク質が、Gタンパク質共役型受容体、キナーゼ、リバーゼ、トランスポーター、プロテアーゼ、イオンチャンネルのいずれかである上記（1）から上記（3）のいずれかに記載の結合分子未知タンパク質の結合分子予測方法。

10 （5）結合分子決定残基位置を1又は2以上特定する工程において、シークエンスアラインメントを構成するアミノ酸残基と結合分子の種類とから結合分子決定残基位置を1又は2以上特定する上記（1）から（4）のいずれかに記載の結合分子未知タンパク質の結合分子予測方法。

15 （6）式1、又は式2のいずれか又は両方を用いて結合分子決定残基位置を決定する上記（1）から（4）のいずれか1項に記載の結合分子未知タンパク質の結合分子予測方法。

（7）式3、式4、式5のいずれかひとつ以上を用いて結合分子決定残基位置を決定する上記（1）から（4）のいずれか1項に記載の結合分子未知タンパク質の結合分子予測方法。

20 （8）リガンド決定残基—リガンド分類情報を得る工程が、リガンド既知タンパク質のアミノ酸残基のうち、関数f3(n)の値が一番小さなりガンド決定残基位置にあるものを抽出する工程と、リガンド決定残基—リガンド分類情報にあげられたりガンド既知タンパク質のうち、抽出されたリガンド決定残基と一致するものの数（A）を求める工程と、リガンド決定残基—リガンド分類情報にあげられたりガンド既知タンパク質のうち抽出されたりガンド決定残基と一致するもののうちで、リガンド又はリガンドの種類が当該リガンド既知タンパク質のものと一致する数（B）を求める工程と、リガンド既知タンパク質のアミノ酸残基のうち関数f3(n)の値が二番目に小さい又はx番目（ここで、xは2より大きく100より小さな整数を表す。）に小さいリガンド決定残基位置にあるものを抽出する工程と、リガンド決定残基—リガンド分類情報にあげられたりガンド既知タンパク質

のうち、抽出されたリガンド決定残基と一致するものの数（C）を求める工程と、リガンド決定残基—リガンド分類情報にあげられたりガンド既知タンパク質のうち抽出されたリガンド決定残基と一致するもののうちで、リガンド又はリガンドの種類が当該リガンド既知タンパク質のものと一致する数（D）を求める工程
5 と、（A）と（C）との和（E）を求める工程と、（B）と（D）との和（F）を求める工程とを含み、（E）と（F）を更に表示するリガンド決定残基—リガンド分類情報を得る上記（7）に記載の結合分子未知タンパク質の結合分子予測方法。

（9）アミノ酸配列と結合分子とが既知である少なくとも2以上の結合分子既
10 知タンパク質について、当該結合分子既知タンパク質のシークエンスアラインメントと、結合分子又は結合分子の種類とを対応付けた結合分子既知タンパク質分類情報を得る工程と、当該結合分子既知タンパク質分類情報を用いて、結合分子既知タンパク質のシークエンスアラインメントのうち結合分子を決定することに
15 関与すると想定される位置である結合分子決定残基位置を1又は2以上特定する工程と、当該結合分子決定残基位置におけるアミノ酸残基（結合分子決定残基）と、結合分子、又は結合分子の種類とを対応付けることにより、結合分子決定残基と結合分子との相関関係を表す結合分子決定残基—結合分子分類情報を得る工
程とを含む結合分子未知タンパク質の結合分子予測方法。

（10）結合分子決定残基と結合分子との相関関係を表す結合分子決定残基—
20 結合分子分類情報に、前記結合分子既知タンパク質と同じ種類の結合分子未知タンパク質について前記結合分子既知タンパク質間のシークエンスアラインメントに対して結合分子未知タンパク質の配列を整列させて得られた結合分子未知タンパク質のシークエンスアラインメントのうち結合分子決定残基に関する情報を入力し、当該結合分子未知タンパク質に結合する結合分子、又は結合分子の種類を
25 予測する結合分子未知タンパク質の結合分子予測方法。

この方法によれば、結合分子が未知であるタンパク質のアミノ酸配列、及び／又はアミノ酸配列を用いて得られるシークエンスアラインメントに関する情報を得るだけで結合分子又は結合分子の種類を予測することが可能となる。これにより、従来の3次元構造まで予測するような分子モデリング法に比べ格段に迅速か

つ低成本に結合分子（リガンド等）を予測することができる。更に、本発明によれば様々な種類の結合分子が未知であるタンパク質に対してその結合分子又は結合分子の種類を予測することができる。また、結合分子が未知であるタンパク質に実際にあらゆる結合分子の候補が結合するかどうか実験するよりも容易かつ迅速に結合分子又は結合分子の種類を予測することができる。結合分子決定残基－結合分子分類情報を得ることによって、結合分子未知タンパク質のシークエンスアラインメントを得るのみで当該情報を当該表に当てはめ、当該結合分子未知タンパク質に結合する結合分子又は結合分子の種類を容易に予測することが可能となる。

10 また、上記課題のうち少なくとも一つは以下の発明によって解決される。すなわち、

(1 1) 上記（1）～（1 0）のいずれかに記載した結合分子未知タンパク質の結合分子予測方法を用いて、結合分子未知タンパク質に結合する結合分子、又は結合分子の種類を予測する工程を含む医薬の製造方法、

15 (1 2) 医薬が、中枢疾患、炎症性疾患、循環器疾患、癌、代謝性疾患、免疫系疾患または消化器系疾患の予防剤、又は治療剤のいずれか又は両方である上記（1 1）に記載の医薬の製造方法である。

また、上記課題の少なくとも一つは以下の発明によって解決される。すなわち

20 (1 3) 式 6 又は下記式 7 のいずれか又は両方を用いた結合分子決定残基位置を決定する方法。

(1 4) 式 8 を用いた結合分子決定残基位置を決定する方法、

25 (1 5) 結合分子既知タンパク質のアミノ酸配列又はシークエンスアラインメントと、結合分子又は結合分子の種類に関する情報とを用いて、結合分子既知タンパク質のシークエンスアラインメントの位置のうち結合分子を決定することに関与すると想定される位置（結合分子決定残基位置）におけるアミノ酸残基である結合分子決定残基と、結合分子または結合分子の種類との相関関係を表す結合分子決定残基－結合分子分類情報を得る、結合分子未知タンパク質の結合分子を予測するためのコンピュータであって、当該コンピュータは、結合分子既知タン

パク質のシークエンスアラインメントに関する情報を入力するシークエンスアラインメント入力手段と、前記シークエンスアラインメント入力手段により入力された結合分子既知タンパク質のアミノ酸配列又はシークエンスアラインメントと、結合分子又は結合分子の種類に関する情報を記憶するシークエンスアラインメント結合分子記憶手段と、前記シークエンスアラインメント結合分子記憶手段により記憶された結合分子既知タンパク質のアミノ酸配列又はシークエンスアラインメントと、結合分子又は結合分子の種類に関する情報を用いて前記結合分子決定残基位置を決定する結合分子決定残基位置決定手段と、前記結合分子決定残基位置におけるアミノ酸残基（結合分子決定残基）と、結合分子又は結合分子の種類とを対応付けることにより、結合分子決定残基と結合分子または結合分子の種類との相関関係を表す結合分子決定残基—結合分子分類情報を得る結合分子決定残基—結合分子分類情報取得手段と、前記結合分子既知タンパク質と同じ種類の結合分子未知タンパク質について前記結合分子既知タンパク質間のシークエンスアラインメントに対して結合分子未知タンパク質の配列を整列させて得られた結合分子未知タンパク質のシークエンスアラインメントに関する情報を入力するシークエンスアラインメント入力手段とを具備し、結合分子決定残基—結合分子分類情報に、シークエンスアラインメント入力手段により入力された結合分子未知タンパク質のシークエンスアラインメントに関する情報を用いて、当該結合分子未知タンパク質の結合分子、又は結合分子の種類を予測する、結合分子未知タンパク質の結合分子を予測するためのコンピュータ、

（16）前記結合分子決定残基位置決定手段が、少なくとも式9又は式10のいずれか又は両方の関数を用いる上記（14）に記載の結合分子未知タンパク質の結合分子を予測するためのコンピュータ、

（17）前記結合分子決定残基位置決定手段が、式9で表される関数を用いる上記（15）又は上記（16）に記載の結合分子未知タンパク質の結合分子を予測するためのコンピュータ、

（18）結合分子未知タンパク質の結合分子を予測するためのコンピュータであって、当該結合分子未知タンパク質と同じ種類であり結合する結合分子が既知である結合分子既知タンパク質のシークエンスアラインメントのうち当該結合分

子既知タンパク質に結合する分子を決定することに関与すると想定される位置である結合分子決定残基位置と、当該結合分子決定残基位置における結合分子既知タンパク質のアミノ酸残基である結合分子決定残基と、当該結合分子決定残基に対応した結合分子既知タンパク質の結合分子又は結合分子の種類に関する情報
5 を記憶した記憶手段と、前記結合分子既知タンパク質と同じ種類の結合分子未知タンパク質について前記結合分子既知タンパク質間のシークエンスアラインメントに対して結合分子未知タンパク質の配列を整列させて得られた結合分子未知タンパク質のシークエンスアラインメントに関する情報を入力するシークエンスアラインメント入力手段と、入力されたシークエンスアラインメントに関する情報
10 と記憶手段に記憶される情報とから当該結合分子未知タンパク質の結合分子又は結合分子の種類を決定する結合分子決定手段と、決定された結合分子未知タンパク質に結合する結合分子又は結合分子の種類を表示する表示手段とを具備し、シークエンスアラインメント入力手段により入力された結合分子未知タンパク質のシークエンスアラインメントに関する情報と、記憶手段に記憶された結合分子決定残基と当該結合分子決定残基に対応した結合分子既知タンパク質の結合分子又は結合分子の種類に関する情報とに基づいて結合分子決定手段により結合分子未知タンパク質の結合分子又は結合分子の種類を予測し、結合分子決定手段により予測された当該結合分子未知タンパク質の結合分子又は結合分子の種類を表示手段により表示する結合分子未知タンパク質の結合分子を予測するためのコンピュ
15 ータである。このようなコンピュータによれば、結合分子既知タンパク質のシークエンスアラインメントに基づいて結合分子決定残基－結合分子分類情報を得ることができ、これにより結合分子未知タンパク質のシークエンスアラインメントを得るのみでその結合分子又は結合分子の種類を容易に予測することができるこ
20 となる。

25 さらに本発明は、

(19) コンピュータを、結合分子既知タンパク質のシークエンスアラインメントに関する情報を入力するシークエンスアラインメント入力手段と、前記シークエンスアラインメント入力手段により入力された結合分子既知タンパク質のアミノ酸配列又はシークエンスアラインメントと、結合分子又は結合分子の種類に

関する情報を記憶するシークエンスアラインメント結合分子記憶手段と、前記シークエンスアラインメント結合分子記憶手段により記憶された結合分子既知タンパク質のアミノ酸配列又はシークエンスアラインメントと、結合分子又は結合分子の種類に関する情報を用いて前記結合分子決定残基位置を決定する結合分子
5 決定残基位置決定手段と、前記結合分子決定残基位置におけるアミノ酸残基（結合分子決定残基）と、結合分子又は結合分子の種類とを対応付けることにより、結合分子決定残基と結合分子または結合分子の種類との相関関係を表す結合分子決定残基－結合分子分類情報を得る結合分子決定残基－結合分子分類情報取得手段と、前記結合分子既知タンパク質と同じ種類の結合分子未知タンパク質について
10 前記結合分子既知タンパク質間のシークエンスアラインメントに対して結合分子未知タンパク質の配列を整列させて得られた結合分子未知タンパク質のシークエンスアラインメントに関する情報を入力するシークエンスアラインメント入力手段と、して機能させるプログラム、

(20) 前記結合分子決定残基位置決定手段が、少なくとも式12又は式13
15 のいずれか又は両方の関数を用いる上記(19)に記載のプログラム、

(21) 前記結合分子決定残基位置決定手段が、式14で表される関数を用いる上記(19)又は(20)に記載のプログラム、

(22) コンピュータを、結合分子未知タンパク質と同じ種類であり結合する結合分子が既知である結合分子既知タンパク質のシークエンスアラインメントのうち当該結合分子既知タンパク質に結合する分子を決定することに関与すると想定される位置である結合分子決定残基位置と、当該結合分子決定残基位置における結合分子既知タンパク質のアミノ酸残基である結合分子決定残基と、当該結合分子決定残基に対応した結合分子既知タンパク質の結合分子又は結合分子の種類とに関する情報を記憶した記憶手段と、前記結合分子既知タンパク質と同じ種類の結合分子未知タンパク質について前記結合分子既知タンパク質間のシークエンスアラインメントに対して結合分子未知タンパク質の配列を整列させて得られた結合分子未知タンパク質のシークエンスアラインメントに関する情報を入力するシークエンスアラインメント入力手段と、入力されたシークエンスアラインメントに関する情報と記憶手段に記憶される情報とから当該結合分子未知タンパク質

の結合分子又は結合分子の種類を決定する結合分子決定手段と、決定された結合分子未知タンパク質に結合する結合分子又は結合分子の種類を表示する表示手段として機能させるプログラム、および、

（23）上記（19）～（22）のいずれか1項に記載のプログラムを記憶した記録媒体、等を提供する。

図面の簡単な説明

図1は、本発明のリガンド決定残基－リガンド分類情報作成までの工程表を表す。

図2は、本発明のリガンド決定残基位置特定工程の一態様を示す工程表である。

図3は、本発明のリガンド決定残基位置特定工程の別の一態様を示す工程表である。

図4は、ウシロドプシンとTGR23-1とのシークエンスアラインメントの結果を表す。

図5は、FLIPRを用いて測定した種々の濃度のヒトTGR23-2リガンド（1-20）によるTGR23-1発現CHO細胞の細胞内Caイオン濃度上昇活性を示す。

図6は、FLIPRを用いて測定した種々の濃度のヒトTGR23-2リガンド（1-20）によるTGR23-2発現CHO細胞の細胞内Caイオン濃度上昇活性を示す。

発明を実施するための最良の形態

本発明は、アミノ酸配列と結合分子とが既知である結合分子既知タンパク質について、少なくとも2以上の結合分子既知タンパク質のシークエンスアラインメントと、結合分子又は結合分子の種類とを対応付けた結合分子既知タンパク質分類情報を得る工程と、前記結合分子既知タンパク質分類情報を用いて、結合分子既知タンパク質のシークエンスアラインメントの位置のうち結合分子を決定することに関与すると想定される位置である結合分子決定残基位置を1又は2以上特

定する工程と、前記結合分子決定残基位置におけるアミノ酸残基（結合分子決定残基）と、結合分子又は結合分子の種類とを対応付けることにより、結合分子決定残基と結合分子又は結合分子の種類との相関関係を表す結合分子決定残基－結合分子分類情報を得る工程と、前記結合分子既知タンパク質と同じ種類の結合分子未知タンパク質について前記結合分子既知タンパク質間のシークエンスアラインメントに対して結合分子未知タンパク質の配列を整列させ、結合分子未知タンパク質のシークエンスアラインメントを得る工程と、前記結合分子未知タンパク質のシークエンスアラインメントのうち結合分子決定残基についての情報を、結合分子決定残基－結合分子分類情報に当てはめ、結合分子未知タンパク質の結合分子又は結合分子の種類を予測する結合分子未知タンパク質の結合分子予測方法に関する。

結合分子既知タンパク質とは、タンパク質であってそれに結合する生体分子が知られているものを意味する。例えば、リガンドが既知であるレセプターなど生体分子が特異的に結合するタンパク質を意味する。

結合分子未知タンパク質とは、結合分子既知タンパク質と同じ種類のタンパク質であって、結合分子が未知のものを意味する。結合分子既知タンパク質と同じ種類の結合分子未知タンパク質について前記結合分子既知タンパク質間のシークエンスアラインメントに対して結合分子未知タンパク質の配列を整列させて得られた結合分子未知タンパク質のシークエンスアラインメントとは、結合分子未知タンパク質のアミノ酸配列と当該結合分子未知タンパク質と同種類のタンパク質のアミノ酸配列の類似性（相同性）を調べるために置換、挿入、欠失を考慮した上で、挿入、欠失に相当する箇所にギャップを入れ、配列全体を並置したものである。

結合分子未知タンパク質及び結合分子既知タンパク質としては、Gタンパク質共役型受容体（G P C R）、キナーゼ、リバーゼ、トランスポーター、プロテアーゼ、イオンチャンネルがあげられる。これらのうちで、Gタンパク質共役型受容体（G P C R）、キナーゼについて本発明が好ましく適用できる。結合分子未知タンパク質及び結合分子既知タンパク質がGタンパク質共役型受容体（G P C R）である場合、結合分子未知タンパク質をオーファンレセプターとも呼ぶ。な

お、前述の通り、結合分子未知タンパク質と結合分子既知タンパク質とは同じ種類である。例えば、結合分子未知タンパク質がG P C Rであれば結合分子既知タンパク質もG P C Rである。

[結合分子既知タンパク質分類情報]

5 結合分子既知タンパク質分類情報とは、少なくとも2以上の結合分子既知タンパク質のシークエンスアラインメントと結合分子（及び／又は、結合分子の種類）とを対応付けた表を意味する。この表は、紙面のみならず、電子的に保存され視覚により表として認識することができるような態様のものであれば特に限定されるものではない。また、結合分子既知タンパク質のシークエンスアラインメント
10 と結合分子又は結合分子の種類との対応が認識できるものであれば特に限定されるものではない。

[結合分子決定残基位置]

結合分子決定残基位置とは、結合分子既知タンパク質のシークエンスアラインメントの位置であって、結合分子を決定することに関与すると想定される位置を意味する。
15

リガンド決定残基位置の数としては、1以上であれば特に限定されるものではなく、1以上10以下であれば好ましく、2以上6以下であればより好ましく、2であれば特に好ましい。生体を構成するアミノ酸残基の種類は20種しか存在しない。したがって、1つのリガンド決定残基のみで対応付けることのできるリガンドは20種類までである。一方、例えば、G P C Rでは、100種類以上のリガンドが知られているので、リガンド決定残基位置が1つであれば、全てのリガンドとリガンド決定残基とを対応付けることができない。したがって、G P C Rのようにリガンドが20種類以上ある系において、結合分子未知タンパク質のリガンドを予測するには、リガンド決定残基位置が2つ以上あることが望ましい。また、リガンド決定残基位置の数が多いほど予測精度が高まる。リガンドを数種類に分類し（ $X_1 \sim X_p : p$ は分類の数である。）、結合分子未知タンパク質のリガンドがいずれの分類に属するかを決定する場合、リガンドの数（pの値）が少なければ、1つのリガンド決定残基位置のアミノ酸残基のみでリガンドの種類を予測することが可能となる。しかし、かかる場合であっても、一般に2つ以上のリガンド決定残基位置のアミノ

酸残基を組合せてリガンドの種類を予測した方が予測精度は高くなる。

[結合分子決定残基]

結合分子決定残基とは、前記結合分子決定残基位置におけるアミノ酸残基を意味する。また、結合分子既知タンパク質について複数種類の結合分子の特定残基位置（1つ又は2つ以上）と複数種類のアミノ酸残基とを組合せ、結合分子決定残基としても良い。例えば、シークエンスアラインメント第2番目と第8番目のアミノ酸残基位置を結合分子決定残基位置の一例とし、シークエンスアラインメント第9番目と第11番目のアミノ酸残基を他の例とする場合である。このように異なった結合分子決定残基位置を組合せることにより、結合分子又は結合分子の種類の予測精度を向上させることができるとなるからである。結合分子決定残基に関する情報とは、前記結合分子決定残基位置におけるアミノ酸残基に関する情報を意味する。例えば、タンパク質のシークエンスアラインメントの第2番目と第8番目が結合分子決定残基位置であれば、シークエンスアラインメントの第2番目と第8番目が、結合分子決定残基の位置に関する情報である。そして、シークエンスアラインメントのうち、第2番目と第8番目のアミノ酸残基の種類についての情報と当該結合分子決定残基の位置に関する情報をあわせて結合分子決定残基に関する情報となる。

[結合分子決定残基－結合分子分類情報]

結合分子決定残基－結合分子分類情報とは、結合分子決定残基と結合分子または結合分子の種類との相関関係を表す表である。この表は、紙面のみならず、電子的に保存され視覚により表として認識することができるような態様のものであれば特に限定されるものではない。また、結合分子決定残基と結合分子又は結合分子の種類との対応が認識できるようなものであれば特に限定されるものではない。

25 [結合分子]

結合分子としては、生体高分子である結合分子既知タンパク質及び結合分子未知タンパク質に結合しうるものであれば特に限定されるものではないが、例えば、レセプタータンパク質に結合するリガンドや、G P C R に結合する G α タンパク質などがあげられる。

[結合分子の種類]

結合分子の種類とは、同じ種類の結合分子既知タンパク質に複数の結合分子が存在する場合にそれらをその機能や性質などに応じて分類したものである。例えば、G P C R のリガンドを、モノアミン、脂質、ペプチドに分類する場合があげられる。

[コンピュータ]

本発明のコンピュータは、一定の計算等をすることができる電子的デバイスであれば、特に限定されるものではなく、たとえば、パーソナルコンピューター、スーパーコンピュータ、モバイル等公知のコンピュータであってもよい。ブラウザを搭載したコンピュータであれば、インターネットに接続することができ、公知の W e b (ウェブ) サイトにアクセスすることができるので特に好ましい。

以下、結合分子がリガンドである場合を例にしてリガンド決定残基ーリガンド分類情報作成までの工程を説明する。

[リガンド決定残基ーリガンド分類情報作成までの工程]

図 1 は、リガンド決定残基ーリガンド分類情報作成までの工程の一例を表し、以下の工程からなる。すなわち、アミノ酸配列とりガンド（及び／又はリガンドの種類）とが既知である少なくとも 2 以上の結合分子既知タンパク質についてのシークエンスアラインメントとりガンド（及び／又はリガンドの種類）に関する情報を取得する工程 (S101) 、シークエンスアラインメントとりガンド（及び／又はリガンドの種類）を対応付け、結合分子既知タンパク質分類情報を得るシークエンスアラインメントリガンド分類情報取得工程 (S102) 、シークエンスアラインメントリガンド分類情報取得工程により得られる結合分子既知タンパク質分類情報を用いて結合分子既知タンパク質のシークエンスアラインメントのうちリガンド（及び／又はリガンドの種類）を決定することに関与すると想定される位置であるリガンド決定残基位置を 1 又は 2 以上特定するリガンド決定残基位置特定工程 (S103) 、リガンド決定残基位置特定工程により特定されたリガンド決定残基位置におけるアミノ酸残基（リガンド決定残基）によって、リガンド（及び／又はリガンドの種類）を対応付けたりガンド決定残基とりガンド（及び／又はリガンドの種類）との相関関係を表すリガンド決定残基ーリガンド分類情報を得るリ

ガンド決定残基ーリガンド分類工程（S104）である。以下、各工程について説明する。なお、以降リガンドの種類を含めて単にリガンドという場合もある。

[ステップ101]

まず、リガンド決定残基ーリガンド分類情報を作成するために、アミノ酸配列

5 とりガンドとが既知である少なくとも2以上の結合分子既知タンパク質についてのシークエンスアラインメントとリガンド（及び／又はリガンドの種類）に関する情報を取得する（S101）。この工程では、複数の結合分子既知タンパク質についてアミノ酸配列とリガンド（及び／又はリガンドの種類）に関する情報を取得するが、アミノ酸配列からシークエンスアラインメントを求めるても良いし、複数
10 の結合分子既知タンパク質について既にシークエンスアラインメントが求められているれば、そのシークエンスアラインメントに関する情報を直接取得しても良い。結合分子既知タンパク質についてのシークエンスアラインメントとリガンド（及び／又はリガンドの種類）に関する情報を取得する方法は、特に限定されるものではなく、データベースから当該情報を取得しても、計算により当該情報を取得しても良い。データベースとしては、少なくとも100種類以上の結合分子既知タンパク質についてのシークエンスアラインメントとリガンド（及び／又はリガンドの種類）に関する情報を収録しているデータベースが好ましく、500以上収録していればより好ましく、1000以上収録していれば特に好ましい。結合分子既知タンパク質の数が多いほど、リガンド決定残基ーリガンド分類情報の精度が高まる
15 からである。

20 公知のデータベースとしては、結合分子既知タンパク質のシークエンスアラインメントとリガンド（及び／又はリガンドの種類）を記述しているものであれば特に限定されるものではなく、例えばG P C R D B (<http://www.GPCR.org/7tm/>) があげられる。シークエンスアラインメントは、公知の計算方法によって求めることもできる。シークエンスアラインメントに関して公知の計算法としては、例えばClustal WやBLASTが挙げられるが、手動で計算しても良い。また、リガンドに関する分類をあらかじめ作成しておき、リガンドに関する情報が入力されれば、リガンドの種類が自動的に求められるようにしてき、リガンドの情報を入手すれば、リガンドの種類に関する情報も得られるようにしてよい。

[ステップ102]

次に、ステップ101で取得された2以上の結合分子既知タンパク質についてのシークエンスアラインメントトリガンド（及び／又はリガンドの種類）に関する情報について、シークエンスアラインメントトリガンド（及び／又はリガンドの種類）に関する情報を対応付け、結合分子既知タンパク質分類情報を得る（シークエンスアラインメントトリガンド分類情報取得工程：S102）。ステップ101において、2以上の結合分子既知タンパク質についてのシークエンスアラインメントトリガンド（及び／又はリガンドの種類）に関する情報をデータベース等から取得する際に、すでにシークエンスアラインメントトリガンド（及び／又はリガンドの種類）が対応付けられていれば、シークエンスアラインメントトリガンド（及び／又はリガンドの種類）が対応付けられたままの情報を取得してもよい。

[ステップ103]

次に、ステップ102のシークエンスアラインメントトリガンド分類情報取得工程により得られる結合分子既知タンパク質分類情報を用いて、結合分子既知タンパク質のシークエンスアラインメントのうちリガンド（及び／又はリガンドの種類）を決定することに関与すると想定される位置であるリガンド決定残基位置を1又は2以上特定する（リガンド決定残基位置特定工程：S103）。

この工程においては、対象とするタンパク質の種類や性質に合わせ好ましい閾数を組合せてリガンド決定残基位置を1又は2以上特定する。この工程の好ましい実施態様については、後述する。

[ステップ104]

次に、リガンド決定残基位置特定工程により特定されたリガンド決定残基位置におけるアミノ酸残基（リガンド決定残基）によって、リガンド（及び／又はリガンドの種類）を対応付けたリガンド決定残基トリガンドとの相関関係を表すリガンド決定残基－リガンド分類情報を得る（リガンド決定残基－リガンド分類工程：S104）。

この工程では、それぞれのG P C Rについてリガンド決定残基位置、リガンド決定残基、及びリガンド（及び／又はリガンドの種類）に関する情報を、シークエンスアラインメントトリガンド分類情報から取得し、リガンド決定残基－リガ

ド分類情報を得る。

このようなリガンド決定残基－リガンド分類情報を用いれば、リガンド未知のG P C Rについても、そのシークエンスアラインメントを求めるだけで、そのリガンド（及び／又はリガンドの種類）を予測することが可能となる。

5 [ステップ103の一実施態様]

図2に従って、リガンド決定残基位置特定工程（S103）の好ましい一態様を説明する。この工程は、リガンド決定残基位置の数が1つの場合である。リガンドをp種類に分け、それぞれX1からXpに分類する。ステップ102で得られた結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質の全てのシークエンスアラインメントとリガンド（及び／又はリガンドの種類）に関する情報を下記の式15に入力し、関数f1(n)の値が小さいものから、リガンド決定残基位置の候補とする（リガンド決定残基位置候補選択工程：S201）。このとき、得られたりガンド決定残基位置の候補におけるシークエンスアラインメントについて、シークエンスアラインメント不能（-で表される）が、全G P C Rの3%以上存在すれば、当該リガンド決定残基位置をリガンド決定残基位置として採用しないことが好ましい。

$$f1(n) = \sum_{Res} (N(Res, Xq) \times N(Res, Xr)) \dots \dots \text{式15}$$

20

[式15中、nは、f1(n)が、結合分子既知タンパク質のシークエンスアラインメントのうち第n番目のアミノ酸残基についての評価関数であることを表し、Resは、アミノ酸残基の種類を表し、Xq及びXrは、リガンドを表し、qは1からp-1までの整数を表し、rはqより大きくp以下である整数を表し、pはリガンドの数を表し、N(Res, Xq)は、結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のうち、シークエンスアラインメントのn番目のアミノ酸残基がResであり、かつリガンドがXqであるものの数を表し、N(Res, Xr)は、結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のうち、シークエンスアラインメントのn番目のアミノ酸残基がResであり、かつリガンドがXrであるものの数を

表す。]

リガンド決定残基位置候補選択工程においてあげられるリガンド決定残基位置候補の数としては、特に限定されるものではないが、1以上100以下が好ましく、1
5 以上10以下であればより好ましく、1以上5以下であれば特に好ましい。リガンド決定残基位置候補が多すぎると後のリガンド決定残基位置候補の信頼性確認工程
が困難になるからである。

ステップ201のリガンド決定残基位置候補選択工程において選択されたリガンド決定残基位置候補は、そのままリガンド決定残基位置としてもよいが、リガンド決定残基位置の候補の信頼性を式2を用いて検討することがより好ましい。（リガンド決定残基位置候補の信頼性検討工程：S202）。この場合、下記式16を用いてリガンド決定残基位置候補の信頼性を検討する。得られたf2(n)の値が小さいほどリガンド決定残基位置としてふさわしいこととなる。

$$15 \quad f2(n) = \sum_{Res} (N(Res, X1) \times N(Res, X2) \cdots \times N(Res, Xp)) \quad \dots \dots \text{式16}$$

[式16中、nは、f2(n)が、結合分子既知タンパク質のシークエンスアラインメントのうち第n番目のアミノ酸残基についての評価関数であることを表し、Res
20 は、アミノ酸残基の種類を表し、X1からXpは、リガンド又はリガンドの種類を表し、pは、リガンド又はリガンドの種類の数を表し、N(Res, X)は、ステップ102
で得られた結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のうち、シークエンスアラインメントのn番目のアミノ酸残基がResであり、かつ
リガンドがXであるものの数を表す。]

25 ステップ201で得られたf1(n)の値及び／またはステップ202で得られたf2(n)の値を用いてリガンド決定残基位置を特定する（リガンド決定残基位置特定工程：S
203）。得られたf2(n)の値が最も小さいものをリガンド決定残基位置としてもよ
いし、f1(n)とf2(n)の値の積が最も小さなものをリガンド決定残基位置としても

よいし、 $f1(n)$ と $f2(n)$ の値の和が最も小さなものをリガンド決定残基位置としてもよい。また、ステップ102で得られた結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のアミノ酸残基位置のうち、 $f1(n)$ が低い方から順位付けをし、更に $f2(n)$ についても同様に順位付けをし、両方の順位を掛け合わせ、最も低いものをリガンド決定残基位置としてもよい。また、得られたりガンド決定残基位置におけるシークエンスアラインメントについてシークエンスアラインメント不能 (-で表される) が、全 G P C R の 3 % 以上存在すれば、当該リガンド決定残基位置をリガンド決定残基位置として採用しないことが好ましい。

[ステップ103の別の実施態様]

図3に従って、リガンド決定残基位置特定工程 (S103) の好ましい別の実施態様を説明する。この工程は、リガンド決定残基位置の数が2つの場合である。まず、ステップ102で得られた結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質の全てのシークエンスアラインメントと、リガンド又はリガンドの種類に関する情報を式1に入力し、リガンド決定残基位置の候補をあげる (リガンド決定残基位置候補選択工程:S301)。

リガンド決定残基位置候補選択工程においてあげられるリガンド決定残基位置候補の数としては、特に限定されるものではないが、1以上100以下が好ましく、1以上20以下であればより好ましく、2以上10以下であれば更に好ましく、2以上6以下であれば特に好ましい。リガンド決定残基位置候補が多いと後のリガンド決定残基位置候補の信頼性確認工程が困難になり、候補が1だと多様なリガンド種類に対応できないからである。

ステップ301のリガンド決定残基位置候補選択工程においてあげられたりガンド決定残基位置候補について、リガンド決定残基位置の候補の信頼性を式2を用いて検討することを含めることはより好ましい実施態様である (リガンド決定残基位置候補の信頼性検討工程:S302)。なお、ステップ302であるリガンド決定残基位置信頼性検討工程を経ない場合は、ステップ301の後ステップ303へと進めばよい。リガンド決定残基位置信頼性検討工程では、少なくともリガンド決定残基位置候補を含むアミノ酸残基位置に関して、式2に結合分子既知タンパク質のシークエンスアラインメントとリガンドに関する情報を入力する。関数 $f2(n)$ の値の小

さいものがリガンド決定位置残基位置としてふさわしい。

関数 $f_2(n)$ の値の小さいものからリガンド決定残基位置として選択してもよいし、 $f_1(n)$ と $f_2(n)$ の値の積が最も小さなものをリガンド決定残基位置として選択してもよいし、 $f_1(n)$ と $f_2(n)$ の値の和が最も小さなものをリガンド決定残基位置として選択してもよい。また、ステップ102で得られた結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のアミノ酸残基位置のうち、 $f_1(n)$ が低い方から順位付けをし、更に $f_2(n)$ についても同様に順位付けをし、両方の順位を掛け合わせ低いものからリガンド決定残基位置として選択してもよい。また、得られたりガンド決定残基位置におけるシークエンスアラインメントについてシークエンスアラインメント不能 (-で表される) が、結合分子既知タンパク質分類情報に記載されたG P C R の 3 %以上存在すれば、当該リガンド決定残基位置をリガンド決定残基位置として採用しないことが好ましい。

次に、ステップ301又はステップ302においてあげられた、リガンド決定残基位置の候補を2つ組合せ、リガンド決定残基位置のペア候補をあげる（リガンド決定残基位置のペア選択工程：S303）。リガンド決定残基位置のペア候補としては、全てのリガンド決定残基位置の候補からなる組合せをあげても良いし、 $f_1(n)$ の好みのリガンド決定残基位置とその他の残基位置との組合せでリガンド決定残基位置のペア候補をあげても良い。

リガンド決定残基位置のペア候補についての情報を式17に入力することによりリガンド決定残基位置のペアを特定する（リガンド決定残基位置のペア特定工程：S304）。

$$f_3(m, n) = \{ (\text{アミノ酸残基ペア種類数}) / w_X + w_1 \times (\text{2交差残基ペア種類数}) \\ + w_2 \times (\text{3交差残基ペア種類数}) + \dots w_{p-1} \times (\text{p交差残基ペア種類数}) + w_A \times (\text{アライメント不能アミノ酸残基数}) + w_B \times (\text{アライメント不能アミノ酸残基ペア数}) \} \dots \quad \text{式17}$$

[式17中、 (m, n) は、 $f_3(m, n)$ が結合分子既知タンパク質のシークエンスアライメントのうち第m番目と第n番目のアミノ酸残基についての評価関数である

ことを表し、アミノ酸残基ペア種類数は、結合分子既知タンパク質のシーケンスアラインメントのうち第m番目と第n番目のアミノ酸残基の組合せの種類の数を表し、2交差残基ペア種類数及び3交差残基ペア種類数はそれぞれ、結合分子既知タンパク質のシーケンスアラインメントのうち第m番目と第n番目のアミノ酸残基の組合せのうちリガンドが2種類及び3種類のものの数を意味し、p交差残基ペア種類数は、結合分子既知タンパク質のシーケンスアラインメントのうち第m番目と第n番目のアミノ酸残基の組合せのうちリガンドがp種類のものの数を意味し、シーケンスアラインメント不能アミノ酸残基数とは、結合分子既知タンパク質のシーケンスアラインメントのうち第m番目と第n番目のアミノ酸残基のうち一方が、好ましい相同意を得るためにシーケンスアラインメント不可能とされた数を意味し、シーケンスアラインメント不能アミノ酸残基ペア数とは、結合分子既知タンパク質のシーケンスアラインメントのうち第m番目と第n番目のアミノ酸残基の両方が、好ましい相同意を得るためにシーケンスアラインメント不可能とされた数を意味し、wXは正の定数、またはアミノ酸ペア種類数を変数とする分布関数であって、アミノ酸ペア種類数が400以下の正の数であるときに最大値を与える分布関数を意味し、w₁…w_{p-1}、wA、wBは、ウエイトであり、正の数である。]

wXは正の定数でもよく、ガウス関数や分布関数でもよい。正の定数の場合、特に限定されるものではないが、1が望ましい。wXが、分布関数の場合、wXは正の定数、またはアミノ酸ペア種類数を変数とする分布関数であって、アミノ酸ペア種類数が400以下であるときに最大値を与える分布関数を意味し、特に限定されるものではない。例えば、アミノ酸ペア種類数を変数とするガウス関数、ローレンツ関数などでもよい。その最大値は、正の数であれば特に限定されるものではない。この分布関数の最大値を与えるような変数の値としては、400以下が好ましく、20～200であればより好ましく、40～140であれば更に好ましい。この値は、リガンド決定アミノ酸ペアの総数が $20 \times 20 = 400$ であり、経験上40～140種のアミノ酸ペアによる予測が最も良好な結果を与えることから決定されたものである。分布関数の半値全幅としては、10以上100以下が好ましく、20以上7

0以下であればより好ましく、30以上50以下であれば更に好ましい。分布関数の最大値としては、他のウエイトとの関係にもよるが、1が好ましい。ウエイト ($w_1 \dots w_{p-1}, w_A, w_B$) の値は、正の数であれば特に限定されるものではない。しかし、例えば、3交差残基ペアが存在する場合、2交差残基ペアよりもリガンドを予測する上で好ましくないため w_2 の値が w_1 の値よりも大きいことが好ましい。例えば、リガンドの数が3つの場合は、3交差残基ペア以降は存在し得ない。この場合のウエイトの組合せとしては、 w_1 が2で w_2 が5で w_A と w_B が1の組合せがあげられる。

リガンド決定残基位置のペア候補のうち小さな $f_3(m, n)$ の値を与えるものが好ましいリガンド決定残基位置のペアである。

なお、特に明示しないが、リガンド決定残基位置のペアを任意に組合せたりガンド決定残基—リガンド分類情報を得ることは、本発明の好ましい別の実施態様である。組合せ方としては、特に限定されるものではないが、前述の関数 $f_3(n)$ の値が一番小さなものと、二番目及び／又は x 番目（ここで、 x は2より大きく100より小さな整数を表す。）に小さなものを組合せる、関数 $f_3(m, n)$ の値が x 番目に小さなものと y 番目（ここで、 y は、 x とは異なり、2より大きく100より小さな整数を表す。）に小さなものを組合せる等があげられる。例えば、 $f_3(n)$ 関数の値が、29番目までのリガンド決定残基を任意に組合せることができる。ここで、 x 番目及び y 番目の x 及び y は、コンピュータにあらかじめ入力されてもよいし、ユーザからの入力情報をコンピュータが受け取り、リガンド決定残基の組合せを作成してもよい。 $f_3(m, n)$ 関数の値が最小なものと2番目に小さなものを組合せたのみでは、リガンド及び／又はリガンドの種類を予測することができるリガンド位置決定残基は限られたものとなる。しかし、このように、複数のリガンド位置決定残基を組合せることで、多くの種類のリガンド未知タンパク質のリガンド及び／又はリガンドの種類を予測することが可能となり、しかも当該予測の精度も高まることとなりうる。このような、リガンド決定残基の組合せは、リガンド決定残基組合せ手段により自動的に組合せることが可能となっていることが、本発明の好ましい実施の態様である。かかる手段を用いれば、リガンド決定残基位置から導かれる予測を補完することができる。なお、上記 x と y

は、それぞれ2より大きく、100より小さいことが好ましく、50より小さければより好ましく、30より小さければ更に好ましく、20より小さければ特に好ましい。

この実施態様では、例えば、あるリガンド既知タンパク質のアミノ酸残基のうち、関数f3(m, n)の値が一番小さなりガンド決定残基位置にあるものを抽出する。そして、リガンド決定残基—リガンド分類情報にあげられたりガンド既知タンパク質のうち、抽出されたリガンド決定残基と一致するものの数を求める。さらに、これらのうちで、リガンド又はリガンドの種類が、当該あるリガンド既知タンパク質と一致するものの数を求める。このような本明細書ではこのようにしてえられた値をN：((リガンド決定残基—リガンド分類情報にあげられたりガンド既知タンパク質のうち、抽出されたリガンド決定残基と一致するものの数) / (リガンド又はリガンドの種類が、当該あるリガンド既知タンパク質と一致するものの数))とする。そして、関数f3(m, n)の値が二番目に小さなりガンド決定残基位置についても同様にNを求める。その後、関数f3(m, n)の値第一番目及び第二番目に小さなりガンド決定残基位置のNの分母同士及び分子同士を足し合わせる。このようにして、リガンド決定残基位置のペアを任意に組合せたりガンド決定残基—リガンド分類情報を得るのである。

以上説明した各工程は、手動によって行われてもよいが所定の媒体又はプログラムをインストールしたコンピュータによって行われることが特に好ましい。また、このコンピュータは、インターネットに接続されており、外部のデータベースにアクセス可能となっていることが特に好ましい。

このようなコンピュータは、少なくとも、結合分子未知タンパク質のシークエンスアラインメントに関する情報を入力するシークエンスアラインメント入力手段と、前記シークエンスアラインメント入力手段により入力された結合分子既知タンパク質のアミノ酸配列又はシークエンスアラインメントと結合分子又は結合分子の種類に関する情報を記憶するシークエンスアラインメント結合分子記憶手段と、前記シークエンスアラインメント結合分子記憶手段により記憶された結合分子既知タンパク質のアミノ酸配列又はシークエンスアラインメントと結合分子又は結合分子の種類に関する情報を用いて前記結合分子決定残基位置を予測す

る結合分子決定残基位置決定手段と、前記結合分子決定残基位置におけるアミノ酸残基（結合分子決定残基）と結合分子または結合分子の種類とを対応付けることにより、結合分子決定残基と結合分子または結合分子の種類との相関関係を表す結合分子決定残基－結合分子分類情報を得る結合分子決定残基－結合分子分類情報取得手段とを具備する。

結合分子決定残基位置決定手段においては、結合分子既知タンパク質のシークエンスアラインメントや結合分子又は結合分子の種類に関する情報を先述した $f_1(n)$ 、 $f_2(n)$ 、 $f_3(m, n)$ 関数のいずれか又はこれらを任意に組合せ結合分子決定残基位置を決定する。

リガンド決定残基－リガンド分類情報は、独自のデータベースを構成していてもよいし、ステップ102で得られる結合分子既知タンパク質分類情報に基づくりレーションナルデータベースとして構成されていてもよい。リガンド決定残基－リガンド分類情報が、結合分子既知タンパク質分類情報に基づくりレーションナルデータベースとして構成されていれば、結合分子未知タンパク質のシークエンスアラインメントとそのリガンドが確認された場合、それを結合分子既知タンパク質分類情報に組込んで、リガンド決定残基位置を容易に再評価することができるため好ましい。

このようなコンピュータであれば、結合分子決定残基－結合分子分類情報に結合分子未知タンパク質のシークエンスアラインメントを入力することにより結合分子又は結合分子の種類を容易に予測することが可能となる。

なお、本発明のコンピュータは、結合分子決定残基－結合分子分類情報をあらかじめインストールしておき、タンパク質のシークエンスアラインメントを入力するシークエンスアラインメント入力手段によって、結合分子未知タンパク質のシークエンスアラインメントを入力することにより、結合分子及び／又は結合分子の種類を予測するコンピュータであってもよい。このようなコンピュータであっても、結合分子決定残基－結合分子分類情報に結合分子未知タンパク質のシークエンスアラインメントを入力することにより結合分子又は結合分子の種類を容易に予測することが可能となる。

[仮想的な結合分子既知タンパク質を用いた例]

以下、仮想的な結合分子既知タンパク質を用いて、本発明のリガンド決定残基一リガンド分類情報作成までの工程及び結合分子未知タンパク質の結合分子種予測工程を説明する。

表1は、仮想的な結合分子既知タンパク質分類情報の例である。この系では、
 5 リガンドが既知でありアミノ酸配列（シークエンスアラインメント）が既知である6つの結合分子既知タンパク質が集められている。そして、リガンドは、P、A
 、Nの3つの分類に分けられている。リガンドの種類に基づいてリガンド決定残基位置を決定する場合は、P、A、Nが、それぞれ式1、式2におけるX1、X2、X3に対応する。また、リガンドに基づいてリガンド決定残基位置を決定する場合は、○×
 10 、××、△△が、それぞれ式1、式2におけるX1、X2、X3に対応する。

表1 仮想的な結合分子既知タンパク質分類表の例

番号	アラインメント	リガンド種類	リガンド
15	1 TLMRK	P	○×
	2 TMMQK	A	××
	3 TCMTK	N	△△
	4 TLLRK	P	○×
	5 TMLQK	A	××
20	6 TLLRA	P	○×

表1において、例えば1番目の結合分子既知たんぱく質のシークエンスアラインメントが、TLMRKであり、その結合分子（リガンド）は○×である。そして、リガンド○×のリガンドの種類は、Pである。

評価関数f1(1)について説明する。上記シークエンスアラインメントのうち1番目のアミノ酸残基はTのみである。したがって、関数f1(n)において、ResはTのみである。

アミノ酸残基がTであって、リガンドがPであるものは3種類あるから、

$N(Res, X1) = N(T, X1) = 3$ となる。同様にして、

30 $N(Res, X2) = N(T, X2) = 2$ 、 $N(Res, X3) = N(T, X3) = 1$ となる。

これから、f1(1)は以下の通りとなる。

$$f1(1) = \sum_{Res} (N(Res, X1) \times N(Res, X2) + N(Res, X1) \times N(Res, X3) + N(Res, X2) \times N($$

$$Res, X3)) = \sum (N(T, X1) \times N(T, X2) + N(T, X1) \times N(T, X3) + N(T, X2) \times N(T, X3)) = 3 \times 2 +$$

5

$$3 \times 1 + 2 \times 1 = 11$$

次に $f2(1)$ について説明する。

$$10 f2(1) = \sum_{Res} (N(Res, X1) \times N(Res, X2) \times N(Res, X3))$$

$$= \sum (N(T, X1) \times N(T, X2) \times N(T, X3)) = 3 \times 2 \times 1 = 6$$

となる。

15 $f1(2)$ と $f2(2)$ について説明する。上記シーケンスアラインメントのうち2番目のアミノ酸残基は L、M、C の3種類である。したがって、関数 $f1(n)$ において、Res は L、M、C である。2番目のアミノ酸残基が L であって、リガンドが P であるものは 3 種類あるから、 $N(L, X1) = N(L, P) = 3$ となる。2番目のアミノ酸残基が L であって、リガンドが A、N のものは存在しないから、 $N(L, X2) = N(L, A) = N(L, X3) = N(L, N) = 0$ となる。2番目のアミノ酸残基が M であって、リガンドが A であるものは 2 種類あるから、 $N(M, X2) = N(M, A) = 2$ となる。2番目のアミノ酸残基が L であって、リガンドが P、N のものは存在しないから、 $N(M, X1) = N(M, P) = N(M, X3) = N(M, N) = 0$ となる。

20 2番目のアミノ酸残基が C であって、リガンドが N であるものは 1 種類あるから、 $N(C, X3) = N(C, N) = 1$ となる。2番目のアミノ酸残基が C であって、リガンドが P、A のものは存在しないから、 $N(C, X1) = N(C, P) = N(C, X2) = N(C, A) = 0$ となる。

25 これから、 $f1(2)$ は以下の通りとなる。

$$f1(2) = \sum_{Res} (N(Res, X1) \times N(Res, X2) + N(Res, X1) \times N(Res, X3) + N(Res, X2) \times$$

$$30 N(Res, X3)) = \sum (N(L, X1) \times N(L, X2) + N(L, X1) \times N(L, X3) + N(L, X2) \times N(L, X3))$$

$$+ \sum (N(M, X1) \times N(M, X2) + N(M, X1) \times N(M, X3) + N(M, X2) \times N(M, X3))$$

29

$$+ \sum (N(C, X1) \times N(C, X2) + N(C, X1) \times N(C, X3) + N(C, X2) \times N(C, X3))$$

$$= \sum (N(L, P) \times N(L, A) + N(L, P) \times N(L, N) + N(L, A) \times N(L, N))$$

5 $+ \sum (N(M, P) \times N(M, A) + N(M, P) \times N(M, N) + N(M, A) \times N(M, N))$

$$+ \sum (N(C, P) \times N(C, A) + N(C, P) \times N(C, N) + N(C, A) \times N(C, N))$$

$$= 0$$

10

また $f_2(2)$ は次の通りとなる。

$$f_2(2) = \sum_{\text{Res}} (N(\text{Res}, X1) \times N(\text{Res}, X2) \times N(\text{Res}, X3))$$

15 $= (N(L, P) \times N(L, A) \times N(L, N)) + (N(M, P) \times N(M, A) \times N(M, N)) + (N(C, P) \times N(C, A) \times N(C, N)) = 0$

同様にして $f_1(3) = 5$ 、 $f_1(4) = 0$ 、 $f_1(5) = 8$ 、 $f_2(3) = 1$ 、 $f_2(4) = 0$ 、 $f_2(5) = 4$ となる。

20

$f_1(n)$ の値が小さなものから並べると、 n が 2、 4、 3、 5、 1 の順番となる。小さな $f_1(n)$ の値を与えるアミノ酸残基位置が、 リガンド決定残基位置の候補である。ここでは、 第 2、 4、 3 番目のアミノ酸残基位置を、 リガンド決定残基位置の候補とする。いくつのアミノ酸残基位置をリガンド決定残基位置の候補とするかはあらかじめ決めておいても良い。これらのアミノ酸残基位置における $f_2(n)$ の値は、 第 1、 5 番目のアミノ酸残基位置における $f_2(n)$ の値に比べ小さいことから、 第 2、 4、 3 番目のアミノ酸残基位置が、 リガンド決定残基位置の候補として望ましい候補であることが確認できる。

[リガンド決定残基位置のペア決定過程]

30 リガンド決定残基位置のペアに対してリガンドが 2 種類あるものを 2 交差残基ペア種とし、 リガンド決定残基位置のペアに対してリガンドが 3 種類あるものを 3 交差残基ペア種とする。第 2、 4、 3 番目のアミノ酸残基位置をそれぞれ組合せ、 リガンド決定残基位置のペア候補をあげる。この例では、 (2, 3)、 (2, 4)、 (3, 4) の 3 種類

のペア候補があげられる。

まず、リガンド決定残基位置のペア候補(2, 3)について検討する。シークエンスアラインメントの2番目と3番目にあるアミノ酸残基の組合せは、(L, M)、(M, M)、(C, M)、(L, L)、(M, L)の5種類である。したがって、「属するアミノ残基ペア種類数」は5である。これら5種類のアミノ酸残基の組合せについて対応するリガンドはそれぞれ一義的に決まるので、2交差残基ペア種と3交差残基ペア種はない。これから $f_3(2, 3)$ の値は5である。同様にして、 $f_3(2, 4)$ の値は、4であり、 $f_3(3, 4)$ の値は、5である。よって残基位置のペア(2, 4)が最も好ましく、リガンド決定残基位置のペアである。

なお、好ましくないアミノ酸残基位置の組合せであれば、 f_3 の値が大きくなることを示すために、アミノ酸残基位置1と5の組合せを用いて $f_3(1, 5)$ を求める。シークエンスアラインメントの1番目と5番目にあるアミノ酸残基の組合せは、(T, K)、(T, A)の2種類である。したがって、「属するアミノ残基ペア種類数」は2である。(T, K)の組合せに対応するリガンドの種類は、P、A、Nの3種であるから、(T, K)は、3交差残基ペア種である。よって、 $f_3(1, 5)$ の値は、7となる。この値は、 $f_3(2, 4)$ の値よりも大きく、関数 f_3 がリガンド決定残基位置のペア決定に適していることが理解できる。

[リガンド決定残基—リガンド分類情報作成工程]

以上よりリガンド決定残基位置のペアがシークエンスアラインメントの(2, 4)番目であることがわかった。これらに対応するアミノ酸残基とそのリガンドを抽出しリガンド決定残基—リガンド分類情報を作成する。

表2 リガンド決定残基—リガンド分類表

(2, 4)	リガンド	リガンドの種類
L, R	○×	P
M, Q	××	A
C, T	△△	N

表2から、例えば、2、4番目のシークエンスアラインメントがそれぞれLと

Rであれば、そのリガンドは○×であり、リガンドの種類はPであると予測される。

[結合分子未知タンパク質の結合分子種予測工程]

以上のようなリガンド決定残基－リガンド分類情報を入手できれば、結合分子未
5 知タンパク質（リガンドが未知の結合分子既知タンパク質）のリガンド決定残基
位置におけるアミノ酸残基を求めるリガンド決定残基－リガンド分類情報に当ては
めることにより、結合分子未知タンパク質のリガンドを予測することが可能とな
る。例えば、ある結合分子未知タンパク質のシークエンスアラインメントを公知
の方法により求め、当該シークエンスアラインメントのうち第2番目と第4番目の
10 アミノ酸残基が、それぞれMとQであれば、その結合分子未知タンパク質のリガン
ドの種類はA、でありリガンドは××（表2）と予想できる。このように、結合分
子未知タンパク質のリガンド（又はリガンドの種類）を予測することにより、そ
のタンパク質とリガンドのペアが決められ、そのリガンドの物理的・化学的・生
物学的性質は容易に推定できることから、そのタンパク質の機能を推定するこ
15 につながる。したがって、本発明の予測方法は新規な医薬を開発する上でも有益
である。

以上の説明では、リガンドについて説明したが、本発明は、リガンドのみなら
ず、結合分子既知タンパク質に結合する分子を予測することができる。例えば、
結合分子既知タンパク質がG P C Rである場合には当該G P C Rに結合する、G
20 α タンパク質を予測することもできる。

各種細胞や臓器における複雑な機能を調節する物質と、その特異的レセプター
タンパク質、特にGタンパク質共役型レセプタータンパク質との関係を明らか
にすることは、各種細胞や臓器における複雑な機能を解明し、それら機能と密接
に関連した医薬品開発に非常に重要な手段を提供することとなる。本発明の結合
25 分子未知タンパク質の結合分子予測方法を用いれば、例えば、G P C Rのリガン
ド及び／又はリガンドの種類を予測することが可能となる。G P C Rのリガンド
及び／又はリガンドの種類を予測することができれば、当該G P C Rの生体内で
の機能を予測することにつながる。そして、例えば、機能が予測され、そのリガ
ンド及び／又はリガンドの種類が予測されたG P C Rに関する情報を用いれば、

容易に当該G P C Rが関与する疾患等の予防薬、治療薬を製造することが可能となる。

G P C Rの機能を考慮すると、特に中枢疾患、炎症性疾患、循環器疾患、癌、代謝性疾患、免疫系疾患または消化器系疾患の予防剤、若しくは治療剤のいずれか又は両方の製造方法に本発明は有効に利用されることとなる。

本明細書の配列表の配列番号は、以下の配列を示す。

[配列番号： 1]

ラットT G R 2 3 – 2 リガンド（1 – 1 8）のアミノ酸配列を示す。

[配列番号： 2]

ラットT G R 2 3 – 2 リガンド（1 – 1 5）のアミノ酸配列を示す。

[配列番号： 3]

ラットT G R 2 3 – 2 リガンド（1 – 1 4）のアミノ酸配列を示す。

[配列番号： 4]

15 以下の参考例 6 におけるP C R 反応で使用したプライマーの塩基配列を示す。

[配列番号： 5]

以下の参考例 6 におけるP C R 反応で使用したプライマーの塩基配列を示す。

[配列番号： 6]

以下の参考例 6 におけるP C R 反応で使用したプライマーの塩基配列を示す。

20 [配列番号： 7]

ヒトT G R 2 3 – 2 リガンド前駆体をコードするc D N Aの塩基配列を示す。

[配列番号： 8]

ヒトT G R 2 3 – 2 リガンド前駆体のアミノ酸配列を示す。

[配列番号： 9]

25 ヒトT G R 2 3 – 2 リガンド（1 – 1 8）のアミノ酸配列を示す。

[配列番号： 1 0]

ヒトT G R 2 3 – 2 リガンド（1 – 1 5）のアミノ酸配列を示す。

[配列番号： 1 1]

ヒトT G R 2 3 – 2 リガンド（1 – 1 4）のアミノ酸配列を示す。

[配列番号：12]

ヒトTGR23-2リガンド（1-20）のアミノ酸配列を示す。

[配列番号：13]

以下の参考例7におけるPCR反応で使用したプライマーの塩基配列を示す。

5 [配列番号：14]

以下の参考例7におけるPCR反応で使用したプライマーの塩基配列を示す。

[配列番号：15]

以下の参考例7におけるPCR反応で使用したプライマーの塩基配列を示す。

[配列番号：16]

10 マウスTGR23-2リガンド前駆体をコードするcDNAの塩基配列を示す

。

[配列番号：17]

マウスTGR23-2リガンド前駆体のアミノ酸配列を示す。

[配列番号：18]

15 マウスTGR23-2リガンド（1-18）のアミノ酸配列を示す。

[配列番号：19]

マウスTGR23-2リガンド（1-15）のアミノ酸配列を示す。

[配列番号：20]

マウスTGR23-2リガンド（1-14）のアミノ酸配列を示す。

20 [配列番号：21]

マウスTGR23-2リガンド（1-20）のアミノ酸配列を示す。

[配列番号：22]

以下の参考例8におけるPCR反応で使用したプライマーの塩基配列を示す。

[配列番号：23]

25 ラットTGR23-2リガンド前駆体の一部をコードするcDNAの塩基配列を示す。

[配列番号：24]

ラットTGR23-2リガンド前駆体の一部のアミノ酸配列を示す。

[配列番号：25]

ラット T G R 2 3 - 2 リガンド (1 - 2 0) のアミノ酸配列を示す。

[配列番号： 2 6]

ヒト T G R 2 3 - 2 リガンド (1 - 1 6) のアミノ酸配列を示す。

[配列番号： 2 7]

5 以下の参考例 9 における P C R 反応で使用したプライマーの塩基配列を示す。

[配列番号： 2 8]

以下の参考例 9 における P C R 反応で使用したプライマーの塩基配列を示す。

[配列番号： 2 9]

以下の参考例 9 における P C R 反応で使用したプライマーの塩基配列を示す。

10 [配列番号： 3 0]

ラット T G R 2 3 - 2 リガンド前駆体をコードする c D N A の塩基配列を示す

。

[配列番号： 3 1]

ラット T G R 2 3 - 2 リガンド前駆体のアミノ酸配列を示す。

15 [配列番号： 3 2]

以下の参考例 1 0 における P C R 反応で使用したプライマー 1 の塩基配列を示す。

[配列番号： 3 3]

以下の参考例 1 0 における P C R 反応で使用したプライマー 2 の塩基配列を示す。

20

[配列番号： 3 4]

以下の参考例 1 1 における T G R 2 3 - 1 発現 C H O 細胞の T G R 2 3 - 1 遺伝子発現量を測定するのに使用したプライマーの塩基配列を示す。

[配列番号： 3 5]

25 以下の参考例 1 1 における T G R 2 3 - 1 発現 C H O 細胞の T G R 2 3 - 1 遺伝子発現量を測定するのに使用したプライマーの塩基配列を示す。

[配列番号： 3 6]

以下の参考例 1 1 における T G R 2 3 - 1 発現 C H O 細胞の T G R 2 3 - 1 遺伝子発現量を測定するのに使用したプローブの塩基配列を示す。 5' 端は 6 - カ

ルボキシフルオレセイン (F a m) で、3' 端は 6 -カルボキシテトラメチルローダミン (T a m r a) で標識されている。

[配列番号： 3 7]

ヒト由来Gタンパク質共役型レセプタータンパク質 T G R 2 3 - 1 (ヒト T G
5 R 2 3 - 1) のアミノ酸配列を示す。

[配列番号： 3 8]

ヒト由来Gタンパク質共役型レセプタータンパク質 T G R 2 3 - 1 をコードす
る c D N A の塩基配列を示す。

[配列番号： 3 9]

10 ヒト由来Gタンパク質共役型レセプタータンパク質 T G R 2 3 - 2 (ヒト T G
R 2 3 - 2) のアミノ酸配列を示す。

[配列番号： 4 0]

ヒト由来Gタンパク質共役型レセプタータンパク質 T G R 2 3 - 2 をコードす
る c D N A の塩基配列を示す。

15

実施例

以下の実施例においては、結合分子既知タンパク質および結合分子未知タンパ
ク質として G P C R をあげるが、本発明は、その要旨を超えない限りこれらに限
定されるものではない。

20

[実施例 1]

(G P C R のリガンドの種類の予測)

G P C R のシークエンスアラインメント及びリガンドが登録されている G P C
R D B (データベース) から、1152種類の G P C R のシークエンスアラインメン
25 トおよびリガンドに関する情報を取得し結合分子既知タンパク質分類情報 (表 3
) を得た。その後、取得したリガンドを3つの種類に分けた。3つの種類は以下の
通りである。

P Peptides (ペプチド) 、 Chemokines (ケモカイン) 、 Glycoproteins (糖
タンパク質)

A Monoamines (モノアミン)、(アドレナリン、アセチルコリン、ドーパ
ミン、セレトニン、ヒスタミン)

N Lipids (脂質)

1152種類のG P C Rのシークエンスアラインメントおよびリガンドに関する情
5 報をコンピュータに入力した。入力された1152種類のG P C Rのシークエンスア
ライメントおよびリガンドに関する情報を $f_1(n)$ 関数及び $f_2(n)$ 関数を用いたリ
ガンド決定残基位置候補選択手段により選択した6種類のリガンド決定残基位置
候補を表3に示す。

10 リガンド決定残基位置候補選択手段においては、 $f_1(n)$ 関数の値と、 $f_2(n)$ 関数
の値の積が小さいものをリガンド残基位置候補として選ぶ。それらのうち、シー
クエンスアラインメント不能(−)が、全体の3%以上存在しているものがあれば、
15 リガンド決定残基位置候補から除いた。このようにして、リガンド残基位置候補
を20個選択した。表3は、それらのうちより好ましい6個のリガンド残基位置
候補について、関数 $f_1(n)$ 、関数 $f_2(n)$ の評価値及び順位をそれぞれ表したもの
である。

表3 好ましい6個のリガンド残基位置候補について、関数 $f_1(n)$ 、関数 $f_2(n)$ の
評価値及び順位をそれぞれ表したもの

20 残基位置	$f_1(n)$		$f_2(n)$	
	順位	評価値	順位	評価値
82	1	9222	10	51554
86	5	13317	3	22593
90	2	10112	1	21168
25 91	13	15016	12	62274
209	18	16389	21	81468
230	10	14618	25	93622

30 表3によれば、例えば、シークエンスアラインメントが86番の残基位置につ
いての、関数 $f_1(n)$ の順位が5位で、関数 $f_2(n)$ の順位は3位であることがわかる。
上記6種類のリガンド決定残基位置候補を組合せ、360個のリガンド決定残基位

置のペア候補をあげた。360個のリガンド決定残基位置のペア候補に関する情報を式3に代入し、リガンド決定残基位置のペアを決定した。その結果、 $f_3(n)$ の値が最も小さく、最も好ましいリガンド決定残基位置（のペア）は、（86, 90）、すなわち、G P C Rのシークエンスアラインメントのうち第86番目と第90番目のアミノ酸残基位置であった。また、（86, 90）から導き出される予測を補完できるかどうかを指標とし、次に好ましいリガンド決定残基位置として、（209, 211）、と（86, 236）を選択した。

結合分子既知タンパク質分類情報からG P C Rのシークエンスアラインメントのうち第86番目と第90番目のアミノ酸残基の種類とリガンドの種類の数を抽出し、リガンド決定残基ーリガンド分類情報を得た。それを抜粋したものを表4に示す。表4から、例えば1152種類のG P C Rのうち、アミノ酸残基位置86番目及び90番目のアミノ酸がそれぞれAとGであるものは、86種類あり、それらのリガンドは全てN（脂質）に分類されることがわかる。このように、殆どのG P C Rは、アミノ酸残基位置86番目及び90番目のアミノ酸によって、そのリガンドを予測することができる。

表4 リガンド決定残基ーリガンド分類表（結合分子決定残基ー結合分子分類表）

	アミノ酸残基		リガンドの種類			合計
	86	90	A	N	P	
20	A	G	0	86	0	86
	D	C	114	0	0	114
	D	M	26	0	0	26
	D	Q	13	0	0	13
25	D	S	68	0	0	68
	D	V	26	0	0	26
	F	L	0	20	6	26
	F	M	0	2	13	15
	G	G	0	58	0	58
30	I	L	0	4	23	27
	I	M	0	0	32	32
	I	V	0	1	7	8
	K	F	0	0	11	11
	K	M	0	0	10	10
35	L	M	0	0	9	9

	M	G	0	26	0	26	
	M	V	11	0	13	24	
	P	M	0	0	7	7	
5	P	V	0	0	8	8	
	Q	I	0	0	15	15	
	Q	M	0	0	25	25	
	Q	V	0	0	48	48	
	T	S	0	0	26	26	
	V	F	0	9	0	9	
10	V	G	0	30	0	30	
	V	L	0	0	19	19	
	V	T	17	0	0	17	
	Y	F	0	0	42	42	
	Y	L	0	0	28	28	

15

表4から、例えば、G P C Rのシークエンスアラインメントのうち第86番目と第90番目のアミノ酸残基が、それぞれAとGであるG P C Rは、86種類あり、それらに対するリガンドは全てN（脂質）に分類されるリガンドであることがわかる。

次に、最近リガンドが発見されたG P C Rを無作為に選択し、上記リガンド決定残基—リガンド分類情報の精度（すなわち、今回のリガンド予測方法の精度）を分析した。その結果を表5に示す。

表5

25	G P C R	アミノ酸残基		N	アミノ酸残基		N	アミノ酸残基		N
		86	90		209	211		86	236	
	GPR2	I	M	32/32	V	F	4/26	I	N	61/66
	GPR5	A	L	1/2	Q	L	6/7	A	H	0/0
	GPR8	D	I	1/1	A	T	1/3	D	N	40/234
30	GPR10	Q	V	48/48	R	L	3/20	Q	S	34/34
	GPR13	F	F	4/4	E	L	10/10	F	H	5/9
	GPR14	D	M	26/26	A	Y	0/0	D	N	40/234
	GPR16	C	M	1/3	E	S	0/0	C	S	0/4
	GPR24	D	G	13/13	Q	S	1/1	D	N	40/234
	GPR28	Y	F	42/42	Q	I	1/1	Y	H	71/71
35	GPR29	Y	L	28/28	S	F	12/19	Y	H	71/71
	GPR54	Q	V	48/48	Q	L	5/6	Q	N	52/52
	GPR74	Q	V	48/48	S	Y	2/2	Q	N	52/52
	APJ	I	M	32/32	Y	L	2/7	I	N	61/66
	RFRP	Q	V	48/48	I	Y	0/0	Q	H	3/3

40

表5について説明する。例えば、GPR2のシークエンスアラインメントのうち、第(86, 90)番目（結合分子決定残基位置）のアミノ酸残基は、それぞれI、Mという配列である。結合分子決定残基－結合分子分類情報にある結合分子既知G P C R 1152種類のうち、第(86, 90)番目のアミノ酸残基が、それぞれI、Mとなるものは32種類あって、それらのうちリガンドの種類がペプチドのものは32種類全てである。このG P R 2は、1152種類の結合分子既知G P C Rに含まれていないが、実際のリガンドの種類はペプチドであり、予測されたリガンドの種類と一致している。

また、GPR2のシークエンスアラインメントのうち、結合分子決定残基位置(209, 211)番目のアミノ酸残基は、それぞれV、Fである。そして、上記1152種のうち、209、211番目のシークエンスアラインメントが、V、Fであるものは、26種類あり、それらのうちリガンドの種類がペプチドのものは4種類である。

表5から結合分子決定残基位置(86, 90)が、3種のうちで最も好ましい結合分子決定残基位置であることがわかる。更に、表5によれば、高い精度をもってリガンドの種類を予測できることがわかる。

次に、結合分子決定残基位置(86, 90)、及び(209, 211)を組合せたリガンド決定残基－リガンド分類情報の精度を分析した。

例えば、表5より、GPR2の結合分子決定残基位置(86, 90)、及び(209, 211)のN(1152種のG P C Rであって、結合分子決定残基が、GPR2と一致したものと、リガンドの種類も同一であったものの数)は、それぞれ32/32、及び4/26である。これらを足し合わせ、36/58とする。

このようにして、表5に記載された全てのG P C Rについて、GPR2の結合分子決定残基位置(86, 90)、及び(209, 211)のNを足し合わせた。この結果、評価対象が増大し、結合分子決定残基位置(86, 90)、又は(209, 211)単独でリガンドの種類を予測した場合に比べ、より精度高くなりガンドの種類を予測できることが確認された。

(G P C Rに結合する結合G α タンパク質の種類の予測方法)

まず、G P C Rの結合分子である結合G α タンパク質をG_i、G_q、G_sの3種類に分類した。これは、T I P S (Trends in pharmacological sciences) の2000

Receptor&Ion channel Nomenclature Supplement に従って結合G α タンパク

5 質を3つに分類したものである。なお、簡単の為に、T I P Sの2000 Receptor & Ion channel Nomenclature Supplement 中、G_{i/o}をG_i、G_{q/11}をG_qとした。また、T I P Sの2000 Receptor&Ion channel Nomenclature Supplementにあげられる結合G α タンパク質のから、2種類以上のG α タンパク質に結合するもの、及び G_{i/a1, 3}、G_{i/a2, 3}を例外として除外した。このようにして、約600種類のG P

10 C R及びそのシークエンスアラインメント、並びにそれに結合する結合G α タンパク質、及びその種類に関する情報を得た。

G P C Rのシークエンスアラインメントと、当該Gタンパク質に結合する結合 G α タンパク質の種類に関する情報を、コンピュータに入力した。入力されたG P C Rのシークエンスアラインメントおよび結合G α タンパク質の種類に関する 15 情報をf₁(n)関数及びf₂(n)関数を用いた結合分子決定残基位置候補選択手段により選択した。その結果、2種類の結合分子決定残基位置のペア(177、178)、及び(82、230)が得られた。

結合分子決定残基位置候補選択手段においては、f₁(n)関数の値と、f₂(n)関数の値の積が小さいものを結合分子残基位置候補として選ぶ。それらのうち、シー 20 クエンスアラインメント不能(−)が、全体の3%以上存在しているものがあれば、リガンド決定残基位置候補から除いた。このようにして、リガンド残基位置候補を選択した。

次に、2種類の結合分子決定残基位置のペア(177、178)、及び(82、230)の結合分子の種類を予測する精度を確認した。

25 複数のG P C Rについて結合するG α タンパク質の種類を文献から入手した。そして、結合G α タンパク質が、どのような手法で得られたものであるかを、Ca influxによる場合、Arachidonic acid release:アラキドン酸の放出による場合、PTX(pertussis toxin sensitive)による場合、環状アデノシン一リン酸による場合に分け、それぞれ、Ca、AA、PTX、cAMPとした。なお、CaだけでG_qの判定をし

ている場合は、結合G α タンパク質が他のものである可能性があるので、排除した。このようにして、G P C Rを選択した。

選択されたG P C R及び、文献から得られた結合G P C Rの種類、当該種類が得られた手法、結合分子決定残基位置のペア（177、178）、及び（82、
5 230）におけるシークエンスアラインメントと、それぞれのペアから予測されるG α タンパク質の種類を表6に示す。

表6

10	GPCR	アミノ酸残基		N	アミノ酸残基		N	実験	予想
		177	178		83	230			
	GPR5	R	S	9/9	N	R	0/0	Gi (PTX)	Gi
	GPR8	L	G	4/4	L	T	0/0	Gi (cAMP)	Gq
	GPR13	K	N	3/3	T	E	32/32	Gi (PTX)	Gi
	GPR14	A	R	1/2	F	T	1/1	Gq (AA)	Gq
15	GPR16	R	F	3/3	H	I	3/3	Gq (Ca)	Gq
	GPR24	I	R	1/1	I	I	9/9	Gi (cAMP)	Gi
	GPR39	S	R	0/0	T	E	32/32	Gq (AA)	Gi
	APJ	G	L	2/2	S	T	0/0	Gi (cAMP)	Gi

20 例え、表6のG P R 5について説明する。G P R 5の結合G α タンパク質の種類はGiであり、その種類は、PTXにより取得されたものである。そして、GPR5の第177番目と第178番目のシークエンスアラインメントは、RとSである。このようなシークエンスアラインメントをもつGPCRは9個あり、それらの共役G α タンパク質の種類は、いずれもGiであることがわかる。上記8種のGPCRのうち、GPR5、GPR13
25 、GPR14、GPR16、GPR24、APJの6種について結合G α タンパク質の予想が的中している。以上より本発明によれば、高い精度をもって結合G α タンパク質を予測することが可能となることがわかる。

[実施例3]

30 (T G R 2 3 リガンドの予測)

2種のS D Mペアを組み合わせて評価することにより、高い精度でリガンドの種類を予測できることが明らかとなった。そこで、本発明の方法を用いて、配列

番号：37および配列番号：39で表されるGタンパク質共役型レセプタータンパク質であるTGR23のアミノ酸配列情報のシークエンスアラインメントを行い、その結果から、TGR23リガンドの予測を行った。ここでは、TGR23-1とウシロドプシンのシークエンスアラインメントの結果を図4に示す。

5 図4に示したとおり、TGR23-1における結合分子決定残基位置（86、90）、（209、211）および（86、236）のアミノ酸は、それぞれ（Q、L）、（D、F）および（Q、N）であった。これらのN（1152種のGPCRであって、結合分子決定残基がTGR23と一致したものと、リガンドの種類も同一であったものの数）は、それぞれ0／0、13／27および52／5
10 2であり、評価GPCR数が広範な（86、90）と（86、236）のSDMペアを組み合わせた評価では、リガンド種類はペプチドであると推定された。

以下の参考例では、実際にTGR23（TGR23-1およびTGR23-2）のリガンドがペプチドであることを示す。

15 [参考例1]

（TGR23-2発現CHO細胞に対して特異的にcAMP産生促進活性を示す活性物質のラット全脳抽出物からの精製）

TGR23-2に特異的なリガンド活性を示す物質を、TGR23-2発現CHO細胞に対するcAMP産生促進活性を指標として、ラット全脳から精製した
20 。

ラット全脳抽出物の高速液体クロマトグラフィー（HPLC）フラクションを以下に述べる方法で調製した。日本チャールズリバー（株）より購入したオス8週齢のウイスターラットの全脳400g（200頭分）を順次摘出直後、25頭ずつ沸騰した蒸留水（300ml）に投じて10分間煮沸した。煮沸後、直ちに25 冰冷し、200頭分を合わせて（2.4L）酢酸180mlを加えて終濃度1.0Mとし、低温下ポリトロン（10,000rpm、2分間）を用いて破碎した。破碎液を遠心（8,000rpm、30分）して上清を取り、沈殿には1.0M酢酸2.4Lを加えて再度ポリトロンによって破碎し、一晩攪拌した後、遠心（8,000rpm、30分）して上清を得た。各遠心で得られた上清は、2倍

量 (4. 8 L) の冷アセトンを4℃でゆっくり滴下した後、1回目の遠心により得られた上清については一晩攪拌し、2回目の遠心により得られた上清については4時間攪拌した。アセトンを加えた抽出液は遠心 (8, 000 r p m, 30分) して沈殿を除き、得られた上清については減圧下エバポレーターにてアセトンを留去した。アセトンを留去した抽出液に等量のジエチルエーテルを加え、分液ロートを使って脂質を含むエーテル層を分離して水層を回収した。エーテル脱脂した抽出液はエバポレーターにて減圧下濃縮しエーテルを完全に除去した。濃縮液をガラス纖維濾紙 (アドバンテック、D P 70 (90 mm φ)) で濾過し、濾液をガラス製カラム (30 φ×240 mm) に充填したODSカラム (ダイソー、Daisogel IR-120-ODS-A 63/210 μm) に付した。カラムを1. 0 M酢酸400 ml 1で洗浄後、0. 1%トリフルオロ酢酸を含む60%アセトニトリル500 ml で溶出した。溶出液を減圧下濃縮して溶媒を留去した後、濃縮液を凍結乾燥した。得られた白色粉末1. 2 gを30 mlの0. 1%トリフルオロ酢酸を含む10%アセトニトリルに溶解し、12. 5 mlずつをODSカラム (東ソー、TSKgel 15 ODS-80Ts (21. 5 φ×300 mm)) を用いた10%から60%の0. 1%トリフルオロ酢酸を含むアセトニトリルの濃度勾配溶出法による分取HPLCに付した。HPLCは2回に分けて行い、溶出液は2分毎に60分画にし、2回分の溶出液をまとめた。各分画を減圧下に濃縮・乾固し、残渣に0. 4 mlのジメチルスルホキシド (DMSO) を添加後ボルテックスミキサー、および超音波洗浄機を用いて完全に溶解した。

上記によって得られたHPLCフラクションのDMSO溶液を参考例3に示した方法に従いTGR23-2発現CHO細胞に投与し、細胞内cAMP産生量の測定を行なった結果、分画番号18、20および22～23に顕著なcAMP産生促進活性が認められた。また同様の試料について公知の方法に従いアラキドン酸代謝物遊離活性を調べた結果、顕著な活性が確認された。

これらの活性は他のレセプター発現細胞では認められなかつたことより、ラット全脳抽出物にTGR23-2に特異的なリガンド活性物質が存在することが示された。得られた3つの活性画分をそれぞれ以下の(a)～(c)の方法によりさらに精製した。また、いずれの活性分画についても、以下に述べる最初の陽イ

オン交換カラムを用いた精製工程において得られたcAMP産生促進活性が認められた分画には、同時にFLIPR（モレキュラーデバイス社）によってレセプター特異的な細胞内カルシウム遊離活性が認められた。そこで、それ以降の精製工程における活性の確認には、FLIPRによる細胞内カルシウム遊離活性を指標として用い、活性を示した分画がcAMP産生促進活性を示すことについては適宜確認した。

5 (a) 分画番号18

分画番号18については、10%アセトニトリルを含む10mMギ酸アンモニウム10mlに溶解し、陽イオン交換カラム（東ソー、TSKgel SP-5PW (2.0mm $\phi \times 150\text{ mm}$)）に付した後、10%アセトニトリルを含む10mMから1.0Mのギ酸アンモニウムの濃度勾配により溶出した。活性はギ酸アンモニウム0.4M付近に回収された。活性分画を凍結乾燥後、0.1%トリフルオロ酢酸を含む10%アセトニトリル0.8mlに溶解し、ODSカラム（東ソー、TSKgel ODS-80Ts (4.6 $\phi \times 250\text{ mm}$)）に付した後、0.1%トリフルオロ酢酸を含む10%から25%のアセトニトリルの濃度勾配により溶出した結果、アセトニトリル13%付近に活性が認められた。得られた活性分画を凍結乾燥後、0.1mlのDMSOで溶解し、さらに0.7mlの0.1%ヘプタフルオロ酪酸を含む10%アセトニトリルを加えてODSカラム（和光純薬、Wakosil-II 3C18HG (2.0mm $\phi \times 150\text{ mm}$)）に付した後、0.1%ヘプタフルオロ酪酸を含む10%から37.5%のアセトニトリルの濃度勾配により溶出し、ピーク毎に手動で分取した。活性はアセトニトリル26%付近に認められた。活性画分には、さらに0.7mlの0.1%を含むトリフルオロ酢酸10%アセトニトリルを加え、ODSカラム（和光純薬、Wakosil-II 3C18HG）に付した後、0.1%トリフルオロ酢酸を含む10%から20%のアセトニトリルの濃度勾配によって溶出し、溶出液はピーク毎に手動で分取した。活性はアセトニトリル11%付近に单一ピークとして得られた。この分画に含まれる活性物質は、以下の参考例5に示すようにして構造決定した。

20 (b) 分画番号20

分画番号20については、10%アセトニトリルを含む10mMギ酸アンモニ

ウム 1.0 m l に溶解し、陽イオン交換カラム（東ソー、TSKgel SP-5PW (2.0 mm $\phi \times 150$ mm)）に付した後、10%アセトニトリルを含む1.0 mMから1.0 Mのギ酸アンモニウムの濃度勾配により溶出した。活性はギ酸アンモニウム0.6 M付近に回収された。活性分画を凍結乾燥後、0.1%トリフルオロ酢酸を含む10%アセトニトリル0.8 m l に溶解し、CNカラム（野村化学、Develosil CN-UG-5 (4.6 mm $\phi \times 250$ mm)）に付した後、0.1%トリフルオロ酢酸を含む10%から25%のアセトニトリルの濃度勾配によって溶出した結果、アセトニトリル12%付近に活性が認められた。得られた活性分画を凍結乾燥後、0.1 m l のDMSOで溶解し、さらに0.7 m l の0.1%トリフルオロ酢酸を含む10%アセトニトリルを加えてODSカラム（和光純薬、Wakosil-II 3C18HG (2.0 mm $\phi \times 150$ mm)）に付した後、0.1%トリフルオロ酢酸を含む10%から20%のアセトニトリルの濃度勾配により溶出し、溶出液はピーク毎に手動で分取した。活性はアセトニトリル15%付近に単一ピークとして得られた。この分画に含まれる活性物質を以下の参考例8-3に示すようにして構造決定した。

(c) 分画番号 22～23

分画番号 22～23については、10%アセトニトリルを含む1.0 mMギ酸アンモニウム 1.0 m l に溶解し、陽イオン交換カラム（東ソー、TSKgel SP-5PW (2.0 mm $\phi \times 150$ mm)）に付した後、10%アセトニトリルを含む1.0 mMから1.0 Mのギ酸アンモニウムの濃度勾配により溶出した。活性はギ酸アンモニウム0.4 M付近に回収された。活性分画を凍結乾燥後、0.1%トリフルオロ酢酸を含む10%アセトニトリル0.8 m l に溶解し、CNカラム（野村化学、Develosil CN-UG-5 (4.6 mm $\phi \times 250$ mm)）に付した後、0.1%トリフルオロ酢酸を含む10%から25%のアセトニトリルの濃度勾配によって溶出した結果、アセトニトリル13%付近に活性が認められた。得られた活性分画を凍結乾燥後、0.1 m l のDMSOで溶解し、さらに0.7 m l の0.1%トリフルオロ酢酸を含む10%アセトニトリルを加えてODSカラム（和光純薬、Wakosil-II 3C18HG (2.0 mm $\phi \times 150$ mm)）に付した後、0.1%トリフルオロ酢酸を含む10%から20%のアセトニトリルの濃度勾配により溶出し、ピー

ク毎に手動で分取した。活性はアセトニトリル 16 %付近に認められた。活性分画には、さらに 0.7 ml の 0.1 %ヘプタフルオロ酸を含む 10 %アセトニトリルを加え、ODSカラム（和光純薬、Wakosil-II 3C18HG）に付した後、0.1 %ヘプタフルオロ酸を含む 10 %から 37.5 %のアセトニトリルの濃度勾配によって溶出し、溶出液はピーク毎に手動で分取した。活性はアセトニトリル 28 %付近に単一ピークとして得られた。この分画に含まれる活性物質は、以下の参考例 4 に示すようにして構造決定した。

[参考例 2]

(ラット全脳抽出物中の TGR 23-2 発現 CHO 細胞に対して特異的に細胞内 cAMP 産生促進活性を示す活性物質のプロナーゼによる失活)

参考例 1 で TGR 23-2 発現 CHO 細胞に対して細胞内 cAMP 産生促進活性を示した HPLC 分画 18、20 および 22～23 を、タンパク質分解酵素であるプロナーゼ (Sigma, protease Type XIV (P5147)) で処理し、活性物質がタンパク性であるか否かを調べた。

上記ラット全脳抽出物 HPLC 活性分画 (分画番号 18、20 および 22～23) 各 4 μl を 0.2 M 酢酸アンモニウム 100 μl に加え、これにプロナーゼ 3 μg を添加して 37 °C で 2 時間インキュベートした後、沸騰水中で 10 分間加熱して添加したプロナーゼを失活させた。これに BSA 0.05 mg および CHAPS 0.05 mg を含む蒸留水 1 ml を加え凍結乾燥した。凍結乾燥した試料を、公知の方法に従い TGR 23-2 発現 CHO 細胞に添加して細胞内 cAMP 産生促進活性を測定した。

その結果、いずれの分画の活性もプロナーゼ処理によって完全に消失した。

従って、ラット全脳抽出物中の TGR 23-2 発現 CHO 細胞に対して細胞内 cAMP 産生促進活性を示す活性物質は、いずれもタンパク質またはペプチドであることが明らかとなった。

[参考例 3]

(ラット全脳抽出物の分画番号 20 から得られた TGR 23-2 発現 CHO 細胞

に対して特異的に cAMP 產生促進活性を示す活性物質のアミノ酸配列の決定)

参考例 2 に示したようにラット全脳抽出物の 3 つの分画に含まれる TGR 23-2 発現 CHO 細胞に対して特異的に cAMP 產生促進活性を示す活性物質は、いずれもタンパク性であることが予想されたので、以下のようにそれについてアミノ酸配列解析を行なった。

参考例 1 に示すようにしてラット全脳抽出物の分画番号 20 から得られた TGR 23-2 発現 CHO 細胞に対して特異的に cAMP 產生促進活性を示す活性物質のアミノ酸配列解析および質量分析を行なった。活性ピークを含む溶出液を用いて Procise 491cLC プロテインシーケンサー（アプライドバイオシステム）によるアミノ末端アミノ酸配列分析を行なったところ、N 末端から 18 残基までに SFRNGVGSGVKKTTSFRRRA（配列番号：1）のアミノ酸配列が得られた。同様の溶出液を用いてナノスプレーイオン源（プロタナ）を装着した Thermo Finnigan LCQ イオントラップ質量分析計（サーモクエスト）による質量分析を行なった結果、配列番号：1 のアミノ酸配列から計算される質量値が得られた（実測値：1954.9、計算値：1954.2）。

これより、ラット全脳抽出物の分画番号 20 から得られた TGR 23-2 発現 CHO 細胞に対して特異的に cAMP 產生促進活性を示す活性物質は、配列番号：1 に示すアミノ酸配列を有するものであると決定された。

20 [参考例 4]

（ラット全脳抽出物の分画番号 22～23 から得られた TGR 23-2 発現 CHO 細胞に対して特異的に cAMP 產生促進活性を示す活性物質のアミノ酸配列の決定）

参考例 1 に示すようにしてラット全脳抽出物の分画番号 22～23 から得られた TGR 23-2 発現 CHO 細胞に対して特異的に cAMP 產生促進活性を示す活性物質のアミノ酸配列解析および質量分析を行なった。活性ピークを含む溶出液を用いて Procise 491cLC プロテインシーケンサー（アプライドバイオシステム）によるアミノ末端アミノ酸配列分析を行なったところ、N 末端から 15 残基までに SFRNGVGSGVKKTTSF（配列番号：2）のアミノ酸配列が得られ

た。同様の溶出液を用いてナノスプレーイオン源（プロタナ）を装着したThermo Finnigan LCQイオントラップ質量分析計（サーモクエスト）による質量分析を行なった結果、配列番号：2のアミノ酸配列から計算される質量値が得られた（実測値：1570.8、計算値：1570.8）。

5 これより、ラット全脳抽出物の分画番号22～23から得られたTGR23-2発現CHO細胞に対して特異的にcAMP産生促進活性を示す活性物質は、配列番号：2に示すアミノ酸配列を有するものであると決定された。

[参考例5]

10 (ラット全脳抽出物の分画番号18から得られたTGR23-2発現CHO細胞に対して特異的にcAMP産生促進活性を示す活性物質のアミノ酸配列の決定)

参考例1に示すようにしてラット全脳抽出物の分画番号18から得られたTGR23-2発現CHO細胞に対して特異的にcAMP産生促進活性を示す活性物質のアミノ酸配列解析および質量分析を行なった。活性ピークを含む溶出液を用いてProcise 491cLCプロテインシーケンサー（アプライドバイオシステム）によるアミノ末端アミノ酸配列分析を行なったところ、N末端から14残基までにSFRNGVGSGVKKTTS（配列番号：3）のアミノ酸配列が得られた。同様の溶出液を用いてナノスプレーイオン源（プロタナ）を装着したThermo Finnigan LCQイオントラップ質量分析計（サーモクエスト）による質量分析を行なった結果、配列番号：3のアミノ酸配列から計算される質量値が得られた（実測値：1424.1、計算値：1423.6）。

これより、ラット全脳抽出物の分画番号18から得られたTGR23-2発現CHO細胞に対して特異的にcAMP産生促進活性を示す活性物質は、配列番号：3に示すアミノ酸配列を有するものであると決定された。

25

[参考例6]

(ヒトTGR23-2リガンド前駆体をコードするcDNAのクローニング)

ラット全脳抽出物から得られたTGR23-2発現CHO細胞に対して特異的にcAMP産生促進活性を示す活性ペプチド（本明細書中、ラットTGR23-

2リガンドと記載することができる)のヒトホモログ(本明細書中、ヒトTGR23-2リガンドと記載することができる)の前駆体をコードするcDNAをクローニングするため、ヒト視床下部由来のcDNAを鑄型としたPCRを行なった。

以下の合成DNAプライマーを用い、ヒト視床下部由来のcDNAを鑄型としてPCR法による増幅を行なった。反応液の組成は、ヒト視床下部Marathon-Ready cDNA(CLONTECH) 0.8 μl、配列番号：4および配列番号：5の合成DNAプライマー各1.0 μM、0.2 mM dNTPs、ExTaq(宝酒造) 0.1 μlおよび酵素に付属のExTaqバッファーで、総反応量は20 μlとした。增幅のためのサイクルはサーマルサイクラー(PE Biosystems)を用い、94℃・300秒の加熱の後、94℃・10秒、55℃・30秒、72℃・30秒のサイクルを35回繰り返し、最後に72℃で5分間保温した。次に、DNase、RNase Freeの蒸留水で50倍希釈したPCR反応液2 μl、配列番号：4および配列番号：6の合成DNAプライマー各1.0 μM、0.2 mM dNTPs、ExTaqポリメラーゼ(宝酒造) 0.1 μlおよび酵素に付属のExTaqバッファーで総反応量を20 μlとし、サーマルサイクラー(PE Biosystems)を用い、94℃・300秒の加熱の後、94℃・10秒、55℃・30秒、72℃・30秒のサイクルを35回繰り返し、最後に72℃で5分間保温した。増幅したDNAを2.0%のアガロースゲル電気泳動により分離した後、バンドの部分をカミソリで切り出し、DNAをQIAquick Gel Extraction Kit(キアゲン)を用いて回収した。このDNAを、pGEM-T Easy Vector System(プロメガ)のプロトコールに従ってpGEM-T Easyベクターへクローニングした。これを大腸菌(Escherichia coli) JM109 competent cell(宝酒造)に導入して形質転換した後、cDNA挿入断片を持つクローンをアンピシリンおよびX-galを含むLB寒天培地で選択し、白色を呈するクローンのみを滅菌したつま楊枝を用いて分離し、形質転換体を得た。個々のクローンをアンピシリンを含むLB培地で一晩培養し、QIAwell 8 Plasmid Kit(キアゲン)を用いてプラスミドDNAを調製した。塩基配列の決定のための反応はBigDye Terminator Cycle Sequencing Ready Reaction Kit(PE Biosystems)を用いて行ない、蛍光式自動シーケンサーを用いて解読し、配列番号：7に示すDNA配列を得た。

配列番号：7で表されるDNAの塩基配列には、配列番号：1、配列番号：2および配列番号：3で表されるラット全脳から得られたラットTGR23-2リガンドのアミノ酸配列に極めて類似したアミノ酸配列をコードするようなフレームが存在したことからヒトTGR23-2リガンドの前駆体あるいはその一部を5コードするcDNAであると推定された。

ヒトTGR23-2リガンドと考えられるアミノ酸配列をコードするようなフレームで配列番号：7から翻訳されるアミノ酸配列の5'上流側にはタンパク質翻訳の開始コドンであると予想されるATGが2ヶ所存在するが、疎水性プロットを行なったところ、より5'上流側のATGから翻訳した場合にのみシグナル10配列と推定される疎水性の高い領域が出現したのでこのATGが開始コドンであると推定した。3'側にはヒトTGR23-2リガンドをコードすると考えられる配列の下流に終止コドンが存在した。以上により推定されたヒトTGR23-2リガンド前駆体のアミノ酸配列を配列番号：8に示す。この配列において、ヒトTGR23-2リガンドに相当すると考えられるアミノ酸配列のN末側には、15通常生理活性ペプチドがその前駆体タンパク質から切り出されるとされるLys-Argの配列(Seidah, N. G. et al., Ann. N. Y. Acad. Sci., 839巻, 9-24頁、1998年)が存在した。一方、C末側には終止コドンが存在したが、配列番号：1で表されるアミノ酸配列を有するラットTGR23-2リガンドに対応する配列との間にさらに2残基が存在した。

これより、ヒトTGR23-2リガンドのアミノ酸配列は、ラット全脳抽出物より得られたラットTGR23-2リガンドのアミノ酸配列；配列番号：1〔ラットTGR23-2リガンド(1-18)〕、配列番号：2〔ラットTGR23-2リガンド(1-15)〕および配列番号：3〔ラットTGR23-2リガンド(1-14)〕にそれぞれ対応する、配列番号：9〔ヒトTGR23-2リガンド(1-18)〕、配列番号：10〔ヒトTGR23-2リガンド(1-15)〕および配列番号：11〔ヒトTGR23-2リガンド(1-14)〕で表されるアミノ酸配列、およびさらに配列番号：9のC末側に2残基延長された配列番号：12で表されるアミノ酸配列〔ヒトTGR23-2リガンド(1-20)〕であると推定された。さらに、ヒトTGR23-2リガンドの配列は、マウス

T G R 2 3 - 2 リガンドおよびラット T G R 2 3 - 2 リガンドの配列と異なり、その配列中に A r g - A r g 配列ではなく G l n - A r g 配列を有することから、配列番号：2 6 に示された 1 6 残基のアミノ酸配列〔ヒト T G R 2 3 - 2 リガンド（1 - 1 6）〕もまたリガンドの配列であると推定された。

5

[参考例 7]

(マウス T G R 2 3 - 2 リガンド前駆体をコードする c D N A のクローニング)

ラット全脳抽出物から得られたラット T G R 2 3 - 2 リガンドのマウスホモログ（本明細書中、マウス T G R 2 3 - 2 リガンドと記載することがある）の前駆体をコードする c D N A をクローニングするため、マウス全脳由来の c D N A を鑄型とした P C R を行なった。

以下の合成D N A プライマーを用い、マウス全脳由来の c D N A を鑄型として P C R 法による増幅を行なった。反応液の組成は、マウス全脳Marathon-Ready c D N A (CLONTECH) 0 . 8 μ l 、配列番号：1 3 および配列番号：1 4 の合成D N A プライマー各 1 . 0 μ M 、 0 . 2 mM d N T P s 、 E x T a q (宝酒造) 0 . 1 μ l および酵素に付属の E x T a q バッファーで、総反応量は 2 0 μ l とした。増幅のためのサイクルはサーマルサイクラー (PE Biosystems) を用い、 9 4 °C ・ 5 分間の加熱の後、 9 4 °C ・ 1 0 秒、 6 5 °C ・ 3 0 秒、 7 2 °C ・ 3 0 秒のサイクルを 3 5 回繰り返し、最後に 7 2 °C で 5 分間保温した。次に、 D N a s e 、 R N a s e F r e e の蒸留水で 1 0 0 倍希釈した P C R 反応液 2 μ l 、配列番号：1 3 および配列番号：1 5 の合成D N A プライマー各 1 . 0 μ M 、 0 . 2 mM d N T P s 、 E x T a q ポリメラーゼ (宝酒造) 0 . 1 μ l および酵素に付属の E x T a q バッファーで総反応量は 2 0 μ l とし、サーマルサイクラー (PE Biosystems) を用い、 9 4 °C ・ 5 分間の加熱の後、 9 4 °C ・ 1 0 秒、 6 0 °C ・ 3 0 秒、 7 2 °C ・ 3 0 秒のサイクルを 3 0 回繰り返し、最後に 7 2 °C で 5 分間保温した。増幅したD N A を 2 . 0 % のアガロースゲル電気泳動により分離した後、約 4 4 0 塩基長のD N A をカミソリで切り出し、D N A を QIAquick Gel Extraction Kit (キアゲン) を用いて回収した。このD N A を、 pGEM-T Easy Vector System (プロメガ) のプロトコールに従って pGEM-T Easyベクターへクローニングした。これ

を大腸菌 (*Escherichia coli*) JM109 competent cell (宝酒造) に導入して形質転換した後、cDNA挿入断片を持つクローンをアンピシリンおよびX-galを含むLB寒天培地で選択し、白色を呈するクローンのみを滅菌したつま楊枝を用いて分離し、形質転換体を得た。個々のクローンをアンピシリンを含むLB培地
5 で一晩培養し、QIAwell 8 Plasmid Kit (キアゲン) を用いてプラスミドDNAを調製した。塩基配列の決定のための反応はBigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems) を用いて行ない、蛍光式自動シーケンサーを用いて解読し、配列番号：16で表されるDNA配列を得た。

配列番号：16で表されるDNAの塩基配列には、配列番号：1、配列番号：
10 2および配列番号：3で表されるラット全脳から得られたラットTGR23-2リガンドのアミノ酸配列に極めて類似したアミノ酸配列をコードするようなフレームが存在したことからマウスTGR23-2リガンドの前駆体あるいはその一部をコードするcDNAであると推定された。

マウスTGR23-2リガンドと考えられるアミノ酸配列をコードするような
15 フレームで配列番号：16から翻訳されるアミノ酸配列の5'上流側にはタンパク質翻訳の開始コドンであると予想されるATGが2ヶ所存在するが、疎水性プロットを行なったところ、より5'上流側のATGから翻訳した場合にのみシグナル配列と推定される疎水性の高い領域が出現したのでこのATGが開始コドンであると推定した。このATGコドンのさらに5'上流側には同じフレームで終止コドンが出現した。3'側にはマウスTGR23-2リガンドをコードすると
20 考えられる配列の下流に終止コドンが存在した。以上により推定されたマウスTGR23-2リガンド前駆体のアミノ酸配列を配列番号：17に示す。この配列において、マウスTGR23-2リガンドに相当すると考えられるアミノ酸配列のN末側には、通常生理活性ペプチドがその前駆体タンパク質から切り出される
25 とされるLys-Argの配列 (Seidah, N. G. et al., Ann. N. Y. Acad. Sci. , 839巻、9-24頁、1998年) が存在した。一方、C末側には終止コドンが存在したが、配列番号：1のラットTGR23-2リガンドに対応する配列との間にさらに2残基が存在した。

これより、マウスTGR23-2リガンドのアミノ酸配列は、ラット全脳抽出

物より得られたラットT G R 2 3 – 2リガンドのアミノ酸配列；配列番号：1〔ラットT G R 2 3 – 2リガンド（1 – 1 8）〕、配列番号：2〔ラットT G R 2 3 – 2リガンド（1 – 1 5）〕および配列番号：3〔ラットT G R 2 3 – 2リガンド（1 – 1 4）〕それぞれに対応する、配列番号：1 8〔マウスT G R 2 3 – 2リガンド（1 – 1 8）〕、配列番号：1 9〔マウスT G R 2 3 – 2リガンド（1 – 1 5）〕および配列番号：2 0〔マウスT G R 2 3 – 2リガンド（1 – 1 4）〕で表されるアミノ酸配列、およびさらに配列番号：1 8のC末側に2残基延長された配列番号：2 1で表されるアミノ酸配列〔マウスT G R 2 3 – 2リガンド（1 – 2 0）〕であると推定された。

10

[参考例8]

(ラットT G R 2 3 – 2リガンド前駆体の一部をコードするc D N Aのクローニング)

ラットT G R 2 3 – 2リガンドの前駆体をコードするc D N Aをクローニングするためラット全脳由来のc D N Aを鑄型としたP C Rを行なった。

以下の合成D N Aプライマーを用い、ラット全脳由来のc D N Aを鑄型としてP C R法による增幅を行なった。反応液の組成は、ラット全脳Marathon-Ready c D N A (CLONTECH) 0 . 8 μ l、配列番号：2 2および配列番号：1 4の合成D N Aプライマー各1 . 0 μ M、0 . 2 mM d N T P s、E x T a q (宝酒造) 0 . 1 μ lおよび酵素に付属のE x T a qバッファーで、総反応量は2 0 μ lとした。増幅のためのサイクルはサーマルサイクラー(PE Biosystems)を用い、9 4 °C・5分間の加熱の後、9 4 °C・1 0秒、6 5 °C・3 0秒、7 2 °C・3 0秒のサイクルを3 5回繰り返し、最後に7 2 °Cで5分間保温した。次に、D N a s e、R N a s e F r e eの蒸留水で1 0 0倍希釈したP C R反応液2 μ l、配列番号：2 2のプライマー1 . 0 μ M、配列番号：1 5の合成D N Aプライマー0 . 2 μ M、0 . 2 mM d N T P s、E x T a qポリメラーゼ(宝酒造) 0 . 1 μ lおよび酵素に付属のE x T a qバッファーで総反応量は2 0 μ lとし、サーマルサイクラー(PE Biosystems)を用い、9 4 °C・5分間の加熱の後、9 4 °C・1 0秒、6 0 °C・3 0秒、7 2 °C・3 0秒のサイクルを3 0回繰り返し、最後に7 2 °Cで5

分間保温した。増幅したDNAを2.0%のアガロースゲル電気泳動により分離した後、約200塩基長のDNAをカミソリで切り出し、DNAをQIAquick Gel Extraction Kit（キアゲン）を用いて回収した。このDNAを、pGEM-T Easy Vector System（プロメガ）のプロトコールに従ってpGEM-T Easyベクターへクローニングした。これを大腸菌（Escherichia coli）JM109 competent cell（宝酒造）に導入して形質転換した後、cDNA挿入断片を持つクローンをアンピシリンおよびX-galを含むLB寒天培地で選択し、白色を呈するクローンのみを滅菌したつま楊枝を用いて分離し、形質転換体を得た。個々のクローンをアンピシリンを含むLB培地で一晩培養し、QIAwell 8 Plasmid Kit（キアゲン）を用いて10 プラスミドDNAを調製した。塩基配列の決定のための反応はBigDye Terminator Cycle Sequencing Ready Reaction Kit（PE Biosystems）を用いて行ない、蛍光式自動シーケンサーを用いて解読し、配列番号：23で表されるDNA配列を得た。

配列番号：23で表されるDNAの塩基配列には、配列番号：1、配列番号：15 2および配列番号：3で表されるラット全脳から得られたラットTGR23-2リガンドのアミノ酸配列をコードするフレームが存在した。このフレームを読み取り枠としてDNA配列を翻訳したところ、配列番号：24で表されるアミノ酸配列が得られた。この配列を参考例7で得られたマウスTGR23-2リガンド前駆体のアミノ酸配列（配列番号：16）と比較することにより、本配列がラットTGR23-2リガンド前駆体の一部であるC末側の54アミノ酸からなる配列に相当することが推定された。3'側にはラットTGR23-2リガンドをコードする配列の下流に終止コドンが存在した。この配列において、ラットTGR23-2リガンドのアミノ酸配列のN末側には、通常生理活性ペプチドがその前駆体タンパク質から切り出されるとされるLys-Argの配列（Seidah, N. G. et al.、Ann. N. Y. Acad. Sci.、839巻、9-24頁、1998年）が存在した。一方、C末側には終止コドンが存在したが、配列番号：1のラットTGR23-2リガンドの配列との間にさらに2残基が存在した。

これより、ラットTGR23-2リガンドのアミノ酸配列は、ラット全脳抽出物より得られた配列番号：1〔ラットTGR23-2リガンド（1-18）〕、

配列番号：2 [ラットT G R 2 3 – 2リガンド（1 – 1 5）] および配列番号：3 [ラットT G R 2 3 – 2リガンド（1 – 1 4）] で表されるアミノ酸配列、およびさらに配列番号：1 のC末側に2残基延長された配列番号：2 5 で表されるアミノ酸配列 [ラットT G R 2 3 – 2リガンド（1 – 2 0）] であると推定された。

5 [参考例9]

(ラットT G R 2 3 – 2リガンド前駆体をコードするc D N Aのクローニング)

ラットT G R 2 3 – 2リガンドの前駆体をコードするc D N Aをクローニング
10 するためラット全脳由来のc D N Aを鑄型としたP C Rを行なった。

以下の合成D N Aプライマーを用い、ラット全脳由来のc D N Aを鑄型として
P C R法による増幅を行なった。反応液の組成は、ラット全脳Marathon-Ready c
DNA (CLONTECH) 0. 8 μ l、配列番号：2 7 および配列番号：2 8 の合成D N A
プライマー各1. 0 μ M、0. 2 mM d N T P s、E x T a q (宝酒造) 0. 1
15 μ l および酵素に付属のE x T a qバッファーで、総反応量は2 0 μ lとした。
増幅のためのサイクルはサーマルサイクラー (PE Biosystems) を用い、9 4 °C・
5分間の加熱の後、9 4 °C・1 0秒、6 5 °C・3 0秒、7 2 °C・3 0秒のサイク
ルを3 5回繰り返し、最後に7 2 °Cで5分間保温した。次に、D N a s e、R n
a s e F r e e の蒸留水で5 0倍希釈したP C R反応液2 μ l、配列番号：2 9
20 のプライマー1. 0 μ M、配列番号：2 8 の合成D N Aプライマー0. 2 μ M、
0. 2 mM d N T P s、E x T a qポリメラーゼ (宝酒造) 0. 1 μ l および酵
素に付属のE x T a qバッファーで総反応量は2 0 μ lとし、サーマルサイクラー
25 (PE Biosystems) を用い、9 4 °C・5分間の加熱の後、9 4 °C・1 0秒、6 5
°C・3 0秒、7 2 °C・3 0秒のサイクルを3 0回繰り返し、最後に7 2 °Cで5分
間保温した。増幅したD N Aを2. 0 %のアガロースゲル電気泳動により分離し
た後、約3 5 0塩基長のD N Aをカミソリで切り出し、D N AをQIAquick Gel E
xtraction Kit (キアゲン) を用いて回収した。このD N Aを、pGEM-T Easy Vec
tor System (プロメガ) のプロトコールに従ってpGEM-T Easyベクターへクローニ
ングした。これを大腸菌 (Escherichia coli) JM109 competent cell (宝酒造)

に導入して形質転換した後、cDNA挿入断片を持つクローンをアンピシリンおよびX-galを含むLB寒天培地で選択し、白色を呈するクローンのみを滅菌したつま楊枝を用いて分離し、形質転換体を得た。個々のクローンをアンピシリンを含むLB培地で一晩培養し、QIAwell 8 Plasmid Kit (キアゲン) を用いてプラスミドDNAを調製した。塩基配列の決定のための反応はBigDye Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems) を用いて行ない、蛍光式自動シーケンサーを用いて解読し、配列番号：30で表されるDNA配列を得た。

配列番号：30で表されるcDNAの塩基配列は、参考例8で得たラットTGR23-2リガンド前駆体の一部をコードするDNA配列（配列番号：23）がさらに5'側に延長された配列であった。本配列を、配列番号：1、配列番号：2または配列番号：3で表されるラット全脳から得られたラットTGR23-2リガンドのアミノ酸配列に一致するアミノ酸配列をコードするようなフレームを読み取り枠として翻訳したところ、5'上流側には、ヒトTGR23-2リガンド前駆体およびマウスTGR23-2リガンド前駆体をコードすると推定されるcDNA（配列番号：7および配列番号：16）に存在するタンパク質翻訳の開始コドンであると予想されるATGに対応する位置に、ATGが1ヶ所存在した。また、このATGコドンのさらに5'上流側には同じフレームで終止コドンが出現した。3'側にはマウスTGR23-2リガンドをコードすると考えられる配列の下流に終止コドンが存在した。これより、配列番号：30で表される配列は、ラットTGR23-2リガンド前駆体をコードするcDNA配列であると推定された。配列番号：30で表されるcDNAの塩基配列から翻訳されるアミノ酸配列を配列番号：31に示す。

25 [参考例10]

(TGR23-1(以下、ヒトTGR23-1を、単にTGR23-1と称することもある)発現CHO細胞の作成)

TGR23-1をコードする、配列番号：38で表される塩基配列を有するDNA断片を含有するプラスミドpTB2173を鋳型とし、SalI認識配列を

付加したプライマー1（配列番号：32）およびS p e I認識配列を付加したプライマー2（配列番号：33）を用いてPCR反応を行った。該反応における反応液の組成は上記プラスミド10ngを鑄型として使用し、Pfu Turbo DNA Polymerase（ストラタジーン社）2.5U、プライマー1（配列番号：32）および5 プライマー2（配列番号：33）を各1.0μM、dNTPsを200μM、および反応液に2X GC Buffer I（宝酒造）を25μl加え、50μlの液量とした。PCR反応は、95℃・60秒の後、95℃・60秒、55℃・60秒、72℃・70秒のサイクルを25回繰り返し、最後に72℃・10分の伸長反応を行った。該PCR反応産物をZero Blunt TOPO PCRクローニングキット（インビト 10 ロジエン社）の処方に従いプラスミドベクターpCR-BluntII-TOP0（インビトロジエン社）へサブクローニングした。これをE. coli TOP10（インビトロジエン社）に導入し、pTB2173に含まれるTGR23-1のcDNAを持つクローンを、カナマイシンを含むLB寒天培地で選択した。ここで得られた、5'側および3'側にS a 1 IおよびS p e Iがそれぞれ認識する配列を付加したTGR 15 23-1が導入されたプラスミドによって形質転換されたE. coliのクローンよりPlasmid Miniprep Kit（バイオラッド社）を用いてプラスミドを調製し、制限酵素S a 1 IおよびS p e Iで切斷してインサート部分を切り出した。インサー 20 T DNAは電気泳動後、アガロースゲルより切り出し、次にGel Extraction Kit（キアゲン社）を用いて回収した。このインサートDNAをS a 1 IおよびS p e Iで切斷した動物細胞発現用ベクタープラスミドpAKKO-111H（Hinuma, S. et al. Biochim. Biophys. Acta, Vol. 1219, pp. 251-259 (1994) 記載のpAKKO1.11Hと同一のベクタープラスミド）に加え、DNA Ligation Kit Ver. 2（宝酒造）を用いてライゲーションを行ない、タンパク質発現用プラスミドpAKKO-TGR23-1を構築した。このpAKKO-TGR23-1で形質転換した 25 E. coli TOP10を培養後、Plasmid Miniprep Kit（バイオラッド社）を用いてpAKKO-TGR23-1のプラスミドDNAを調製した。

ハムスターCHO/dhfr⁻細胞を10%ウシ胎児血清を含むα-MEM培地 (with ribonucleosides and deoxyribonucleosides, GIBCO, Cat. No. 12571) でファルコンディッシュ（径3.5cm）に1×10⁵個播種し、5% CO₂

インキュベーターで37℃一晩培養した。上記発現プラスミドpAKK0-TGR23-1 DN A 2 μgをTransfection Reagent FuGENE 6 (Roche社)を用い、添付説明書記載の方法に従ってトランスフェクトし、18時間培養後、新鮮な増殖培地に交換した。さらに10時間培養を続けたのち、トランスフェクトした細胞をトリプシン-EDTA処理により集め、選択培地(10%透析牛胎児血清を含むα-MEM培地(without ribonucleosides and deoxyribonucleosides、GIBCO、Cat. No. 12561))を用いて平底96穴プレート10枚に播種した。3-4日ごとに選択培地を交換しながら培養を続け、2-3週間後にコロニー状に増殖してきたDHF R⁺細胞クローニングを81個取得した。

10

[参考例11]

(TaqMan PCR法を用いたTGR23-1発現CHO細胞株のTGR23-1発現量の定量)

参考例10で得たTGR23-1発現CHO細胞株81クローニングを、96穴プレートに培養し、RNeasy 96 Kit(キアゲン社)を用いて全RNAを調製した。得られた全RNA 50~200ngをTaqMan Gold RT-PCR Kit(P Eバイオシステムズ社)を用いて、逆転写反応を行なった。得られた全RNA 5~20ng相当の逆転写産物、または後述のようにして作製した標準cDNA、1xUniversal PCR Master Mix(P Eバイオシステムズ社)、配列番号:34で表されるプライマーおよび配列番号:35で表されるプライマー各500nM、および配列番号:36で表されるTaqManプローブ100nMを含む反応混合液25μlについてABI PRISM 7700 Sequence Detector(P Eバイオシステムズ社)を用いてPCRを行なった。PCRは、50℃・2分、95℃・10分で処理後、95℃・15秒、60℃・60秒のサイクルを40回繰り返すことにより行なった。

標準cDNAは、配列番号:40で表される塩基配列を有するDNA断片を含有するプラスミドpTB2174の260nmの吸光度を測定して濃度を算出し、正確なコピー数を算出した後、1mM EDTAを含む10mM Tris-HCl(pH8.0)溶液で希釈し、2コピーから2×10⁶コピーの標準cDNA溶液を調製した。また、TaqMan PCR用プローブおよびプライマーはP

rimers Express (Version1.0) (PEバイオシステムズ社) により設計した。

発現量はABI PRISM 7700 SDSソフトウェアによって算出した。リポーターの蛍光強度が設定された値に達した瞬間のサイクル数を縦軸にとり、標準cDNAの初期濃度の対数値を横軸にとり、標準曲線を作成した。標準曲線より各逆転写産物の初期濃度を算出し、各クローンの全RNA当たりのTGR23-1遺伝子発現量を求めた。その結果、TGR23-1の発現が高かったCHO細胞株11個を選択し24穴プレートに培養した。これらの細胞について、TGR23-1の発現量を再検した。RNeasy Mini Kits (キヤゲン社) を用いて全RNAを調製した後、RNase-free DNase Set (キヤゲン社) を用いてDNase処理をした。得られた全RNAから、上記と同様に逆転写反応し、TaqMan PCR法で各クローンの全RNA当たりのTGR23-1遺伝子発現量を求めた。その結果、TGR23-1発現CHO細胞株クローン49および52が高い発現量を示すことがわかった。

以後の参考例では、これら2つのクローンの発現細胞を用いた。

15

[参考例12]

(ヒトTGR23-2リガンド(1-20) : Ser-Phe-Arg-Asn-Gly-Val-Gly-Thr-Gly-Met-Lys-Lys-Thr-Ser-Phe-Gln-Arg-Ala-Lys-Ser(配列番号:12)の製造)

市販のBoc-Ser(Bzl)-OCH₂-PAM樹脂を、ペプチド合成機ACT90の反応槽に入れ、DCMで膨潤後TFAでBocを除去し、DIEAで中和した。この樹脂をNMPに懸濁し、HOBT-DIPCIでBoc-Lys(Cl-Z)を縮合した。反応後ニンヒドリンテストで遊離のアミノ基の有無を調べ、ニンヒドリンテストがプラスの時には同じアミノ酸を再度縮合した。再縮合後においてもニンヒドリンテストがプラスの時には無水酢酸でアセチル化した。このサイクルを繰り返しBoc-Ala、Boc-Arg(Tos)、Boc-Gln、Boc-Phe、Boc-Ser(Bzl)、Boc-Thr(Bzl)、Boc-Lys(Cl-Z)、Boc-Lys(Cl-Z)、Boc-Met、Boc-Gly、Boc-Thr(Bzl)、Boc-Gly、Boc-Val、Boc-Gly、Boc-Asn、Boc-Arg(Tos)、Boc-Phe、Boc-Ser(Bzl)を配列順に縮合し、所望の保護ペプチド樹脂0.24gを得た。この樹脂をp-クレゾール1.5mlとともにフッ

化水素約 1.5 ml 中、0℃で 60 分攪拌した後フッ化水素を減圧留去し、残留物にジエチルエーテルを加えて濾過した。濾過物に水と酢酸を加えペプチドを抽出し、樹脂と分離した。抽出液を濃縮し 50% 酢酸で充填したセファデックス（商標）G-25 カラム（2.0 × 80 cm）に付し、同溶媒で展開、主要画分を集め凍結乾燥した。その一部（4.5 mg）を LiChroprep（商標）RP-18 を充填した逆相クロマトカラム（2.6 × 60 cm）に付け、0.1% TFA 水 200 ml で洗浄、0.1% TFA 水 300 ml と 0.1% TFA 含有 25% アセトニトリル水 300 ml を用いた線型勾配溶出を行い、主要画分を集め凍結乾燥し目的とするペプチド 1.2.7 mg を得た。

10 ESI-MS：分子量 MW 2188.0 (理論値 2187.5) HPLC 溶出時間 10.6 分

カラム条件：カラム：Wakosil 5C18T 4.6 × 100 mm

溶離液：A 液 - 0.1% TFA 水、B 液 - 0.1% TFA 含有アセトニトリルを用い A/B : 95/5 ~ 45/55 へ直線型濃度勾配溶出 (25 分)

15 流速：1.0 ml / 分

[参考例 13]

(FLIPR を用いたヒト TGR 23-2 リガンド (1-20) による TGR 23-1 発現 CHO 細胞および TGR 23-2 発現 CHO 細胞の細胞内 Ca イオン濃度上昇活性の測定)

参考例 12 で得られたヒト TGR 23-2 リガンド (1-20) を種々の濃度で、公知の方法に従って、TGR 23-1 発現 CHO 細胞および TGR 23-2 発現 CHO 細胞に投与し、細胞内 Ca イオン濃度上昇活性を FLIPR を用いて測定したところ、ヒト TGR 23-2 リガンド (1-20) は、濃度依存的に TGR 23-1 発現 CHO 細胞および TGR 23-2 発現 CHO 細胞の細胞内 Ca イオン濃度上昇を促進した。結果を図 5 および図 6 に示す。

これより、配列番号：12 で表されるアミノ配列を有するポリペプチド [ヒト TGR 23-2 リガンド (1-20)] が、TGR 23-1 および TGR 23-2 対する細胞内 Ca イオン濃度上昇活性を有することが明らかである。

産業上の利用可能性

本発明によれば、結合分子が未知であるタンパク質のアミノ酸配列（及び／又はアミノ酸配列を用いて得られるシークエンスアラインメント）に関する情報を得るだけで結合分子又は結合分子の種類を予測することが可能となる。これにより、従来の3次元構造まで予測する分子モデリング法に比べ格段に迅速に結合分子（リガンド等）を予測することができる。

更に、本発明によれば様々な種類の結合分子が未知であるタンパク質に対してその結合分子又は結合分子の種類を予測することができる。また、結合分子が未知であるタンパク質に実際にあらゆる結合分子が結合するかどうか実験するよりも容易かつ迅速に結合分子又は結合分子の種類を予測することができる。

公知の技術を用いることで、相同性を有したタンパク質グループに共通の機能は、シークエンスアラインメントを計算することや立体構造モデルを作成することで推定できる。一方、シークエンスアラインメントによる類似性評価や分子ドッキング計算から、リガンド（およびその種類）の特定は、現在の技術では不可能である。しかし、医薬品開発に有用な個々のGPCRが有している個別の機能を推定するためには、結合するリガンド・共役Gタンパク質の予測が必要である。このような、GPCR・リガンドのセットが推定・決定されて初めて、共役Gタンパク質を通じた細胞内応答の調査、生体内分布や発現量変化の測定、遺伝子導入・欠損動物の作成などの詳細な機能研究が進展することになる。リガンド決定残基を用いた予測方法およびそのためのコンピュータは、このような医薬品開発に必須な分子同定・機能解明に直接役に立つものである。

本発明によれば、きわめて容易に結合分子未知タンパク質（オーファンGタンパク質共役レセプター等）の結合分子又は結合分子の種類を予測することができ、かかる知見に基づけば、実験を経なくとも当該結合分子未知タンパク質が関与する疾患等の予防薬や治療薬を容易に製造することが可能となる。

請求の範囲

1. 結合分子未知タンパク質に結合する結合分子を予測する結合分子未知タンパク質の結合分子予測方法であって、

5 アミノ酸配列と結合分子とが既知である結合分子既知タンパク質について、少なくとも 2 以上の結合分子既知タンパク質のシークエンスアラインメントと、結合分子又は結合分子の種類とを対応付けた結合分子既知タンパク質分類情報を得る工程と、

前記結合分子既知タンパク質分類情報を用いて、結合分子既知タンパク質のシークエンスアラインメントの位置のうち結合分子を決定することに関与すると想定
10 される位置である結合分子決定残基位置を 1 又は 2 以上特定する工程と、

前記結合分子決定残基位置におけるアミノ酸残基（結合分子決定残基）と、結合分子又は結合分子の種類とを対応付けることにより、結合分子決定残基と結合分子又は結合分子の種類との相関関係を表す結合分子決定残基－結合分子分類情報を得る工程と、

15 前記結合分子既知タンパク質と同じ種類の結合分子未知タンパク質について前記結合分子既知タンパク質間のシークエンスアラインメントに対して結合分子未知タンパク質の配列を整列させ、結合分子未知タンパク質のシークエンスアラインメントを得る工程と、

前記結合分子未知タンパク質のシークエンスアラインメントのうち少なくとも 1
20 種類の結合分子決定残基についての情報を、結合分子決定残基－結合分子分類情報に当てはめ、結合分子未知タンパク質の結合分子又は結合分子の種類を予測する工程とを含む、

結合分子未知タンパク質の結合分子予測方法。

2. 前記結合分子が、リガンド、調節因子、エフェクター、補酵素のいずれかである請求項 1 に記載の結合分子未知タンパク質の結合分子予測方法。

3. 前記結合分子が、2 以上の種類に分類され、当該分類された結合分子の種類を予測する請求項 1 又は 2 に記載の結合分子未知タンパク質の結合分子予測方法。

4. 組合分子未知タンパク質が、G タンパク質共役型受容体、キナーゼ、リバー

ゼ、トランスポーター、プロテアーゼ、イオンチャネルのいずれかである請求項1から3のいずれか1項に記載の結合分子未知タンパク質の結合分子予測方法。

5. 結合分子決定残基位置を1又は2以上特定する工程において、シークエンスアラインメントを構成するアミノ酸残基と結合分子の種類とから結合分子決定残基位置を1又は2以上特定する請求項1から4のいずれか1項に記載の結合分子未知タンパク質の結合分子予測方法。

6. 下記式1、又は下記式2のいずれか又は両方を用いて結合分子決定残基位置を決定する請求項1から4のいずれか1項に記載の結合分子未知タンパク質の結合分子予測方法。

10

$$f1(n) = \sum_{Res} (N(Res, Xq) \times N(Res, Xr)) \dots \dots \text{式 } 1$$

[式1中、nは、 $f1(n)$ が、結合分子既知タンパク質のシークエンスアラインメントのうち第n番目のアミノ酸残基についての評価関数であることを表し、Resは、アミノ酸残基の種類を表し、Xq及びXrは、結合分子又は結合分子の種類を表し、qは1からp-1までの整数を表し、rはqより大きくp以下である整数を表し、pは結合分子又は結合分子の種類の数を表し、 $N(Res, Xq)$ は、結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のうち、シークエンスアラインメントのn番目のアミノ酸残基がResであり、かつ結合分子がXqであるものの数を表し、 $N(Res, Xr)$ は、結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のうち、シークエンスアラインメントのn番目のアミノ酸残基がResであり、かつ結合分子がXrであるものの数を表す。]

25 $f2(n) = \sum_{Res} (N(Res, X1) \times N(Res, X2) \dots \times N(Res, Xp)) \dots \dots \text{式 } 2$

[式2中、nは、 $f2(n)$ が、結合分子既知タンパク質のシークエンスアラインメントのうち第n番目のアミノ酸残基についての評価関数であることを表し、Resは、アミノ酸残基の種類を表し、X1からXpは、結合分子又は結合分子の種類を表し、p

は、結合分子又は結合分子の種類の数を表し、 $N(Res, X)$ は、結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のうち、シークエンスアラインメントのn番目のアミノ酸残基がResであり、かつ結合分子がXであるものの数を表す。]

5 7. 下記式3、下記式4、下記式5のいずれかひとつ以上を用いて結合分子決定
10 残基位置を決定する請求項1から4のいずれか1項に記載の結合分子未知タンパ
ク質の結合分子予測方法。

$$f1(n) = \sum_{Res} (N(Res, Xq) \times N(Res, Xr)) \dots \dots \dots \text{式3}$$

[式3中、nは、 $f1(n)$ が、結合分子既知タンパク質のシークエンスアラインメントのうち第n番目のアミノ酸残基についての評価関数であることを表し、Resは、アミノ酸残基の種類を表し、 Xq 及び Xr は、結合分子又は結合分子の種類を表し、qは1からp-1までの整数を表し、rはqより大きくp以下である整数を表し、pは結合分子又は結合分子の種類の数を表し、 $N(Res, Xq)$ は、結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のうち、シークエンスアラインメントのn番目のアミノ酸残基がResであり、かつ結合分子が Xq であるものの数を表し、 $N(Res, Xr)$ は、結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のうち、シークエンスアラインメントのn番目のアミノ酸残基がResであり、かつ結合分子が Xr であるものの数を表す。]

$$f2(n) = \sum_{Res} (N(Res, X1) \times N(Res, X2) \dots \times N(Res, Xp)) \dots \dots \dots \text{式4}$$

[式4中、nは、 $f2(n)$ が、結合分子既知タンパク質のシークエンスアラインメントのうち第n番目のアミノ酸残基についての評価関数であることを表し、Resは、アミノ酸残基の種類を表し、 $X1$ から Xp は、結合分子又は結合分子の種類を表し、pは、結合分子又は結合分子の種類の数を表し、 $N(Res, X)$ は、結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のうち、シークエンスアライン

メントの n 番目のアミノ酸残基が Res であり、かつ結合分子が X であるものの数を表す。]

$$f_3(m, n) = \{ (\text{アミノ酸残基ペア種類数}) / wX + w1 \times (\text{2交差残基ペア種類数}) \\ + w2 \times (\text{3交差残基ペア種類数}) + \dots w_{p-1} \times (\text{p交差残基ペア種類数}) + wA \times (\text{アライメント不能アミノ酸残基数}) + wB \times (\text{アライメント不能アミノ酸残基ペア数}) \} \dots \dots \dots \text{式5}$$

[式5中、(m, n)は、 $f_3(m, n)$ が結合分子既知タンパク質のシークエンスアライ

10 メントのうち第 m 番目と第 n 番目のアミノ酸残基についての評価関数であることを表し、アミノ酸残基ペア種類数は、結合分子既知タンパク質のシークエンスアライメントのうち第 m 番目と第 n 番目のアミノ酸残基の組合せの種類の数を表し、2交差残基ペア種類数及び3交差残基ペア種類数はそれぞれ、結合分子既知タンパク質のシークエンスアライメントのうち第 m 番目と第 n 番目のアミノ酸残基の組合せのうちリガンドが2種類及び3種類のものの数を意味し、p交差残基ペア種類数は、結合分子既知タンパク質のシークエンスアライメントのうち第 m 番目と第 n 番目のアミノ酸残基の組合せのうちリガンドが p 種類のものの数を意味し、シークエンスアライメント不能アミノ酸残基数とは、結合分子既知タンパク質のシークエンスアライメントのうち第 m 番目と第 n 番目のアミノ酸残基のうち一方が、好ましい相同性を得るためにシークエンスアライメント不可能とされた数を意味し、シークエンスアライメント不能アミノ酸残基ペア数とは、結合分子既知タンパク質のシークエンスアライメントのうち第 m 番目と第 n 番目のアミノ酸残基の両方が、好ましい相同性を得るためにシークエンスアライメント不可能とされた数を意味し、 wX は正の定数、またはアミノ酸ペア種類数を変数とする分布関数であって、アミノ酸ペア種類数が400以下の正の数であるときに最大値を与える分布関数を意味し、 $w1 \dots w_{p-1}$ 、 wA 、 wB は、ウエイトであり、正の数である。]

8. リガンド決定残基—リガンド分類情報を得る工程が、

リガンド既知タンパク質のアミノ酸残基のうち、関数 $f_3(n)$ の値が一番小さなりガ

ンド決定残基位置にあるものを抽出する工程と、

リガンド決定残基—リガンド分類情報にあげられたリガンド既知タンパク質のうち、抽出されたリガンド決定残基と一致するものの数（A）を求める工程と、

リガンド決定残基—リガンド分類情報にあげられたリガンド既知タンパク質のうち抽出されたリガンド決定残基と一致するもののうちで、リガンド又はリガンド

5 の種類が当該リガンド既知タンパク質のものと一致する数（B）を求める工程と、

、

リガンド既知タンパク質のアミノ酸残基のうち閾数 $f_3(n)$ の値が二番目に小さい又は x 番目（ここで、 x は 2 より大きく 100 より小さな整数を表す。）に小さ

10 いリガンド決定残基位置にあるものを抽出する工程と、リガンド決定残基—リガンド分類情報にあげられたリガンド既知タンパク質のうち、抽出されたリガンド決定残基と一致するものの数（C）を求める工程と、リガンド決定残基—リガンド分類情報にあげられたリガンド既知タンパク質のうち抽出されたリガンド決定残基と一致するもののうちで、リガンド又はリガンドの種類が当該リガンド既知

15 タンパク質のものと一致する数（D）を求める工程と、

（A）と（C）との和（E）を求める工程と、

（B）と（D）との和（F）を求める工程と、

を含み、

（E）と（F）を更に表示するリガンド決定残基—リガンド分類情報を得る請求

20 項 7 に記載の結合分子未知タンパク質の結合分子予測方法。

9. アミノ酸配列と結合分子とが既知である少なくとも 2 以上の結合分子既知タンパク質について、当該結合分子既知タンパク質のシークエンスアラインメントと、結合分子又は結合分子の種類とを対応付けた結合分子既知タンパク質分類情報を得る工程と、

25 当該結合分子既知タンパク質分類情報を用いて、結合分子既知タンパク質のシークエンスアラインメントのうち結合分子を決定することに関与すると想定される位置である結合分子決定残基位置を 1 又は 2 以上特定する工程と、

当該結合分子決定残基位置におけるアミノ酸残基（結合分子決定残基）と、結合分子、又は結合分子の種類とを対応付けることにより、結合分子決定残基と結合

分子との相関関係を表す結合分子決定残基－結合分子分類情報を得る工程と、を含む結合分子未知タンパク質の結合分子予測方法。

10. 結合分子決定残基と結合分子との相関関係を表す結合分子決定残基－結合分子分類情報に、前記結合分子既知タンパク質と同じ種類の結合分子未知タンパク質について前記結合分子既知タンパク質間のシークエンスアラインメントに対して結合分子未知タンパク質の配列を整列させて得られた結合分子未知タンパク質のシークエンスアラインメントに関する情報を入力し、当該結合分子未知タンパク質に結合する結合分子、又は結合分子の種類を予測する結合分子未知タンパク質の結合分子予測方法。

10 11. 請求項 1～10 のいずれか 1 項に記載した結合分子未知タンパク質の結合分子予測方法を用いて、結合分子未知タンパク質に結合する結合分子、又は結合分子の種類を予測する工程を含む医薬の製造方法。

12. 医薬が、中枢疾患、炎症性疾患、循環器疾患、癌、代謝性疾患、免疫系疾患または消化器系疾患の予防剤、又は治療剤のいずれか又は両方である請求項 1 15 1 に記載の医薬の製造方法。

13. 下記式 6 又は下記式 7 のいずれか又は両方を用いた結合分子決定残基位置を決定する方法。

$$f1(n) = \sum_{Res} (N(Res, Xq) \times N(Res, Xr)) \dots \dots \text{式 } 6$$

20

[式 6 中、n は、 $f1(n)$ が、結合分子既知タンパク質のシークエンスアラインメントのうち第 n 番目のアミノ酸残基についての評価関数であることを表し、Res は、アミノ酸残基の種類を表し、Xq 及び Xr は、結合分子又は結合分子の種類を表し、q は 1 から p-1 までの整数を表し、r は q より大きく p 以下である整数を表し、p は結合分子又は結合分子の種類の数を表し、 $N(Res, Xq)$ は、結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のうち、シークエンスアラインメントの n 番目のアミノ酸残基が Res であり、かつ結合分子が Xq であるものの数を表し、 $N(Res, Xr)$ は、結合分子既知タンパク質分類情報に存在する結合分子既知タンパ

ク質のうち、シーケンスアラインメントのn番目のアミノ酸残基がResであり、かつ結合分子がXrであるものの数を表す。]

$$f2(n) = \sum_{Res} (N(Res, X1) \times N(Res, X2) \cdots \times N(Res, Xp)) \quad \dots \dots \text{式 7}$$

5

[式 7 中、n は、f2(n) が、結合分子既知タンパク質のシーケンスアラインメントのうち第n番目のアミノ酸残基についての評価関数であることを表し、Res は、アミノ酸残基の種類を表し、X1からXp は、結合分子又は結合分子の種類を表し、p は、結合分子又は結合分子の種類の数を表し、N(Res, X) は、結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のうち、シーケンスアラインメントのn番目のアミノ酸残基がRes であり、かつ結合分子がX であるものの数を表す。]

10

14. 下記式 8 を用いた結合分子決定残基位置を決定する方法。

15

$$f3(m, n) = \{ (\text{アミノ酸残基ペア種類数}) / wX + w1 \times (\text{2交差残基ペア種類数}) \\ + w2 \times (\text{3交差残基ペア種類数}) + \cdots + wp-1 \times (\text{p交差残基ペア種類数}) + wA \times (\text{アラインメント不能アミノ酸残基数}) + wB \times (\text{アラインメント不能アミノ酸残基ペア数}) \} \quad \dots \dots \text{式 8}$$

20

25

[式 8 中、(m, n) は、f3(m, n) が結合分子既知タンパク質のシーケンスアラインメントのうち第m番目と第n番目のアミノ酸残基についての評価関数であることを表し、アミノ酸残基ペア種類数は、結合分子既知タンパク質のシーケンスアラインメントのうち第m番目と第n番目のアミノ酸残基の組合せの種類の数を表し、2交差残基ペア種類数及び3交差残基ペア種類数はそれぞれ、結合分子既知タンパク質のシーケンスアラインメントのうち第m番目と第n番目のアミノ酸残基の組合せのうちリガンドが2種類及び3種類のものの数を意味し、p交差残基ペア種類数は、結合分子既知タンパク質のシーケンスアラインメントのうち第m番目と第n番目のアミノ酸残基の組合せのうちリガンドがp種類のものの数を意

味し、シーケンスアラインメント不能アミノ酸残基数とは、結合分子既知タンパク質のシーケンスアラインメントのうち第m番目と第n番目のアミノ酸残基のうち一方が、好ましい相同性を得るためにシーケンスアラインメント不可能とされた数を意味し、シーケンスアラインメント不能アミノ酸残基ペア数とは
5 、結合分子既知タンパク質のシーケンスアラインメントのうち第m番目と第n番目のアミノ酸残基の両方が、好ましい相同性を得るためにシーケンスアラインメント不可能とされた数を意味し、wXは正の定数、またはアミノ酸ペア種類数を変数とする分布関数であって、アミノ酸ペア種類数が400以下の正の数であるときに最大値を与える分布関数を意味し、w1…wp-1、wA、wBは、ウエイトであり、正の数である。】

15. 結合分子既知タンパク質のアミノ酸配列又はシーケンスアラインメントと、結合分子又は結合分子の種類に関する情報を用いて、結合分子既知タンパク質のシーケンスアラインメントの位置のうち結合分子を決定することに関与すると想定される位置（結合分子決定残基位置）におけるアミノ酸残基である結合分子決定残基と、結合分子または結合分子の種類との相関関係を表す結合分子決定残基—結合分子分類情報を得る、結合分子未知タンパク質の結合分子を予測するためのコンピュータであって、

当該コンピュータは、
結合分子既知タンパク質のシーケンスアラインメントに関する情報を入力する
20 シーケンスアラインメント入力手段と、

前記シーケンスアラインメント入力手段により入力された結合分子既知タンパク質のアミノ酸配列又はシーケンスアラインメントと、結合分子又は結合分子の種類に関する情報を記憶するシーケンスアラインメント結合分子記憶手段と、

25 前記シーケンスアラインメント結合分子記憶手段により記憶された結合分子既知タンパク質のアミノ酸配列又はシーケンスアラインメントと、結合分子又は結合分子の種類に関する情報を用いて前記結合分子決定残基位置を決定する結合分子決定残基位置決定手段と、

前記結合分子決定残基位置におけるアミノ酸残基（結合分子決定残基）と、結合

分子又は結合分子の種類とを対応付けることにより、結合分子決定残基と結合分子または結合分子の種類との相関関係を表す結合分子決定残基—結合分子分類情報を得る結合分子決定残基—結合分子分類情報取得手段と、

前記結合分子既知タンパク質と同じ種類の結合分子未知タンパク質について前記

5 結合分子既知タンパク質間のシークエンスアラインメントに対して結合分子未知タンパク質の配列を整列させて得られた結合分子未知タンパク質のシークエンスアラインメントに関する情報を入力するシークエンスアラインメント入力手段と

、

を具備し、

10 結合分子決定残基—結合分子分類情報に、シークエンスアラインメント入力手段により入力された結合分子未知タンパク質のシークエンスアラインメントに関する情報を用いて、当該結合分子未知タンパク質の結合分子、又は結合分子の種類を予測する、

結合分子未知タンパク質の結合分子を予測するためのコンピュータ。

15 16. 前記結合分子決定残基位置決定手段が、少なくとも下記式9又は式10のいずれか又は両方の関数を用いる請求項15に記載の結合分子未知タンパク質の結合分子を予測するためのコンピュータ。

$$f1(n) = \sum_{Res} (N(Res, Xq) \times N(Res, Xr)) \dots \dots \text{式9}$$

20

[式9中、nは、 $f1(n)$ が、結合分子既知タンパク質のシークエンスアラインメントのうち第n番目のアミノ酸残基についての評価関数であることを表し、Resは、アミノ酸残基の種類を表し、Xq及びXrは、結合分子又は結合分子の種類を表し、qは1からp-1までの整数を表し、rはqより大きくp以下である整数を表し、pは結合分子又は結合分子の種類の数を表し、N(Res, Xq)は、結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のうち、シークエンスアラインメントのn番目のアミノ酸残基がResであり、かつ結合分子がXqであるものの数を表し、N(Res, Xr)は、結合分子既知タンパク質分類情報に存在する結合分子既知タンパ

25

ク質のうち、シーケンスアラインメントのn番目のアミノ酸残基がResであり、かつ結合分子がXrであるものの数を表す。]

$$f_2(n) = \sum_{\text{Res}} (N(\text{Res}, X1) \times N(\text{Res}, X2) \cdots \times N(\text{Res}, Xp)) \quad \dots \quad \text{式 } 10$$

5

[式10中、nは、 $f_2(n)$ が、結合分子既知タンパク質のシーケンスアラインメントのうち第n番目のアミノ酸残基についての評価関数であることを表し、Resは、アミノ酸残基の種類を表し、X1からXpは、結合分子又は結合分子の種類を表し、pは、結合分子又は結合分子の種類の数を表し、 $N(\text{Res}, X)$ は、結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のうち、シーケンスアラインメントのn番目のアミノ酸残基がResであり、かつ結合分子がXであるものの数を表す。]

17. 前記結合分子決定残基位置決定手段が、下記式11で表される関数を用いる請求項15又は16に記載の結合分子未知タンパク質の結合分子を予測するためのコンピュータ。

$$f_3(m, n) = \{ (\text{アミノ酸残基ペア種類数}) / wX + w1 \times (\text{2交差残基ペア種類数}) \\ + w2 \times (\text{3交差残基ペア種類数}) + \cdots w_{p-1} \times (\text{p交差残基ペア種類数}) + wA \times (\text{ア} \\ \text{ラインメント不能アミノ酸残基数}) + wB \times (\text{アラインメント不能アミノ酸残基ペ} \\ \text{ア数}) \} \quad \dots \quad \text{式 } 11$$

[式11中、(m, n)は、 $f_3(m, n)$ が結合分子既知タンパク質のシーケンスアラインメントのうち第m番目と第n番目のアミノ酸残基についての評価関数であることを表し、アミノ酸残基ペア種類数は、結合分子既知タンパク質のシーケンスアラインメントのうち第m番目と第n番目のアミノ酸残基の組合せの種類の数を表し、2交差残基ペア種類数及び3交差残基ペア種類数はそれぞれ、結合分子既知タンパク質のシーケンスアラインメントのうち第m番目と第n番目のアミノ酸残基の組合せのうちリガンドが2種類及び3種類のものの数を意味し、p交差残基

ペア種類数は、結合分子既知タンパク質のシークエンスアラインメントのうち第m番目と第n番目のアミノ酸残基の組合せのうちリガンドがp種類のものの数を意味し、シークエンスアラインメント不能アミノ酸残基数とは、結合分子既知タンパク質のシークエンスアラインメントのうち第m番目と第n番目のアミノ酸残基のうち一方が、好ましい相同性を得るためにシークエンスアラインメント不可能とされた数を意味し、シークエンスアラインメント不能アミノ酸残基ペア数とは、結合分子既知タンパク質のシークエンスアラインメントのうち第m番目と第n番目のアミノ酸残基の両方が、好ましい相同性を得るためにシークエンスアラインメント不可能とされた数を意味し、w_Xは正の定数、またはアミノ酸ペア種類数を変数とする分布関数であって、アミノ酸ペア種類数が400以下の正の数であるときに最大値を与える分布関数を意味し、w₁…w_{p-1}、w_A、w_Bは、ウェイトであり、正の数である。]

18. 結合分子未知タンパク質の結合分子を予測するためのコンピュータであつて、

15 当該結合分子未知タンパク質と同じ種類であり結合する結合分子が既知である結合分子既知タンパク質のシークエンスアラインメントのうち当該結合分子既知タンパク質に結合する分子を決定することに関与すると想定される位置である結合分子決定残基位置と、当該結合分子決定残基位置における結合分子既知タンパク質のアミノ酸残基である結合分子決定残基と、当該結合分子決定残基に対応した結合分子既知タンパク質の結合分子又は結合分子の種類とに関する情報を記憶した記憶手段と、
20

前記結合分子既知タンパク質と同じ種類の結合分子未知タンパク質について前記結合分子既知タンパク質間のシークエンスアラインメントに対して結合分子未知タンパク質の配列を整列させて得られた結合分子未知タンパク質のシークエンスアラインメントに関する情報を入力するシークエンスアラインメント入力手段と

入力されたシークエンスアラインメントに関する情報と記憶手段に記憶される情報とから当該結合分子未知タンパク質の結合分子又は結合分子の種類を決定する結合分子決定手段と、

決定された結合分子未知タンパク質に結合する結合分子又は結合分子の種類を表示する表示手段とを具備し、

シークエンスアラインメント入力手段により入力された結合分子未知タンパク質のシークエンスアラインメントに関する情報と、記憶手段に記憶された結合分子

5 決定残基と当該結合分子決定残基に対応した結合分子既知タンパク質の結合分子又は結合分子の種類に関する情報とに基づいて結合分子決定手段により結合分子未知タンパク質の結合分子又は結合分子の種類を予測し、結合分子決定手段により予測された当該結合分子未知タンパク質の結合分子又は結合分子の種類を表示手段により表示する

10 結合分子未知タンパク質の結合分子を予測するためのコンピュータ。

19. コンピュータを、

結合分子既知タンパク質のシークエンスアラインメントに関する情報を入力するシークエンスアラインメント入力手段と、

15 前記シークエンスアラインメント入力手段により入力された結合分子既知タンパク質のアミノ酸配列又はシークエンスアラインメントと、結合分子又は結合分子の種類に関する情報を記憶するシークエンスアラインメント結合分子記憶手段と、

20 前記シークエンスアラインメント結合分子記憶手段により記憶された結合分子既知タンパク質のアミノ酸配列又はシークエンスアラインメントと、結合分子又は結合分子の種類に関する情報を用いて前記結合分子決定残基位置を決定する結合分子決定残基位置決定手段と、

25 前記結合分子決定残基位置におけるアミノ酸残基（結合分子決定残基）と、結合分子又は結合分子の種類とを対応付けることにより、結合分子決定残基と結合分子または結合分子の種類との相関関係を表す結合分子決定残基—結合分子分類情報を得る結合分子決定残基—結合分子分類情報取得手段と、

前記結合分子既知タンパク質と同じ種類の結合分子未知タンパク質について前記結合分子既知タンパク質間のシークエンスアラインメントに対して結合分子未知タンパク質の配列を整列させて得られた結合分子未知タンパク質のシークエンスアラインメントに関する情報を入力するシークエンスアラインメント入力手段と

して機能させるプログラム。

20. 前記結合分子決定残基位置決定手段が、少なくとも下記式12又は式13のいずれか又は両方の関数を用いる請求項19に記載のプログラム。

5

$$f1(n) = \sum_{Res} (N(Res, Xq) \times N(Res, Xr)) \dots \dots \dots \text{式12}$$

[式12中、nは、 $f1(n)$ が、結合分子既知タンパク質のシーケンスアラインメントのうち第n番目のアミノ酸残基についての評価関数であることを表し、Resは、アミノ酸残基の種類を表し、Xq及びXrは、結合分子又は結合分子の種類を表し、qは1からp-1までの整数を表し、rはqより大きくp以下である整数を表し、pは結合分子又は結合分子の種類の数を表し、 $N(Res, Xq)$ は、結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のうち、シーケンスアラインメントのn番目のアミノ酸残基がResであり、かつ結合分子がXqであるものの数を表し、 $N(Res, Xr)$ は、結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のうち、シーケンスアラインメントのn番目のアミノ酸残基がResであり、かつ結合分子がXrであるものの数を表す。]

20 $f2(n) = \sum_{Res} (N(Res, X1) \times N(Res, X2) \dots \times N(Res, Xp)) \dots \dots \dots \text{式13}$

[式13中、nは、 $f2(n)$ が、結合分子既知タンパク質のシーケンスアラインメントのうち第n番目のアミノ酸残基についての評価関数であることを表し、Resは、アミノ酸残基の種類を表し、X1からXpは、結合分子又は結合分子の種類を表し、pは、結合分子又は結合分子の種類の数を表し、 $N(Res, X)$ は、結合分子既知タンパク質分類情報に存在する結合分子既知タンパク質のうち、シーケンスアラインメントのn番目のアミノ酸残基がResであり、かつ結合分子がXであるものの数を表す。]

30 21. 前記結合分子決定残基位置決定手段が、下記式14で表される関数を用い

る請求項19又は20に記載のプログラム。

$$f_3(m, n) = \{ (アミノ酸残基ペア種類数) / wX + w1 \times (2交差残基ペア種類数) \\ + w2 \times (3交差残基ペア種類数) + \dots wp-1 (p交差残基ペア種類数) + wA \times (ア \\ 5 ラインメント不能アミノ酸残基数) + wB \times (アラインメント不能アミノ酸残基ペ \\ ア数) \} \dots \dots \dots 式14$$

[式14中、(m, n)は、 $f_3(m, n)$ が結合分子既知タンパク質のシークエンスアラインメントのうち第m番目と第n番目のアミノ酸残基についての評価関数であることを表し、アミノ酸残基ペア種類数は、結合分子既知タンパク質のシークエンスアラインメントのうち第m番目と第n番目のアミノ酸残基の組合せの種類の数を表し、2交差残基ペア種類数及び3交差残基ペア種類数はそれぞれ、結合分子既知タンパク質のシークエンスアラインメントのうち第m番目と第n番目のアミノ酸残基の組合せのうちリガンドが2種類及び3種類のものの数を意味し、p交差残基ペア種類数は、結合分子既知タンパク質のシークエンスアラインメントのうち第m番目と第n番目のアミノ酸残基の組合せのうちリガンドがp種類のものの数を意味し、シークエンスアラインメント不能アミノ酸残基数とは、結合分子既知タンパク質のシークエンスアラインメントのうち第m番目と第n番目のアミノ酸残基のうち一方が、好ましい相同性を得るためにシークエンスアラインメント不可能とされた数を意味し、シークエンスアラインメント不能アミノ酸残基ペア数とは、結合分子既知タンパク質のシークエンスアラインメントのうち第m番目と第n番目のアミノ酸残基の両方が、好ましい相同性を得るためにシークエンスアラインメント不可能とされた数を意味し、wXは正の定数、またはアミノ酸ペア種類数を変数とする分布関数であって、アミノ酸ペア種類数が400以下の正の数であるときに最大値を与える分布関数を意味し、w1…wp-1、wA、wBは、ウエイトであり、正の数である。]

22. コンピュータを、

結合分子未知タンパク質と同じ種類であり結合する結合分子が既知である結合分子既知タンパク質のシークエンスアラインメントのうち当該結合分子既知タンパ

ク質に結合する分子を決定することに関与すると想定される位置である結合分子決定残基位置と、当該結合分子決定残基位置における結合分子既知タンパク質のアミノ酸残基である結合分子決定残基と、当該結合分子決定残基に対応した結合分子既知タンパク質の結合分子又は結合分子の種類に関する情報を記憶した記憶手段と、

前記結合分子既知タンパク質と同じ種類の結合分子未知タンパク質について前記結合分子既知タンパク質間のシークエンスアラインメントに対して結合分子未知タンパク質の配列を整列させて得られた結合分子未知タンパク質のシークエンスアラインメントに関する情報を入力するシークエンスアラインメント入力手段と、

入力されたシークエンスアラインメントに関する情報と記憶手段に記憶される情報とから当該結合分子未知タンパク質の結合分子又は結合分子の種類を決定する結合分子決定手段と、

決定された結合分子未知タンパク質に結合する結合分子又は結合分子の種類を表示する表示手段として機能させるプログラム。

23. 請求項19～請求項22のいずれか1項に記載のプログラムを記憶した記録媒体。

図 1

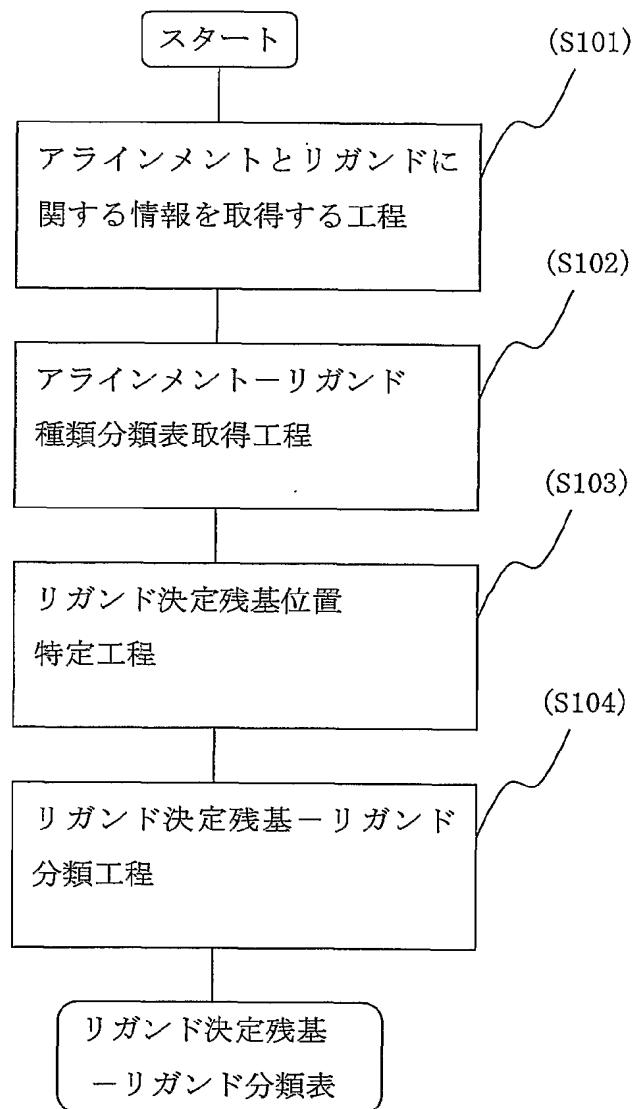


図 2

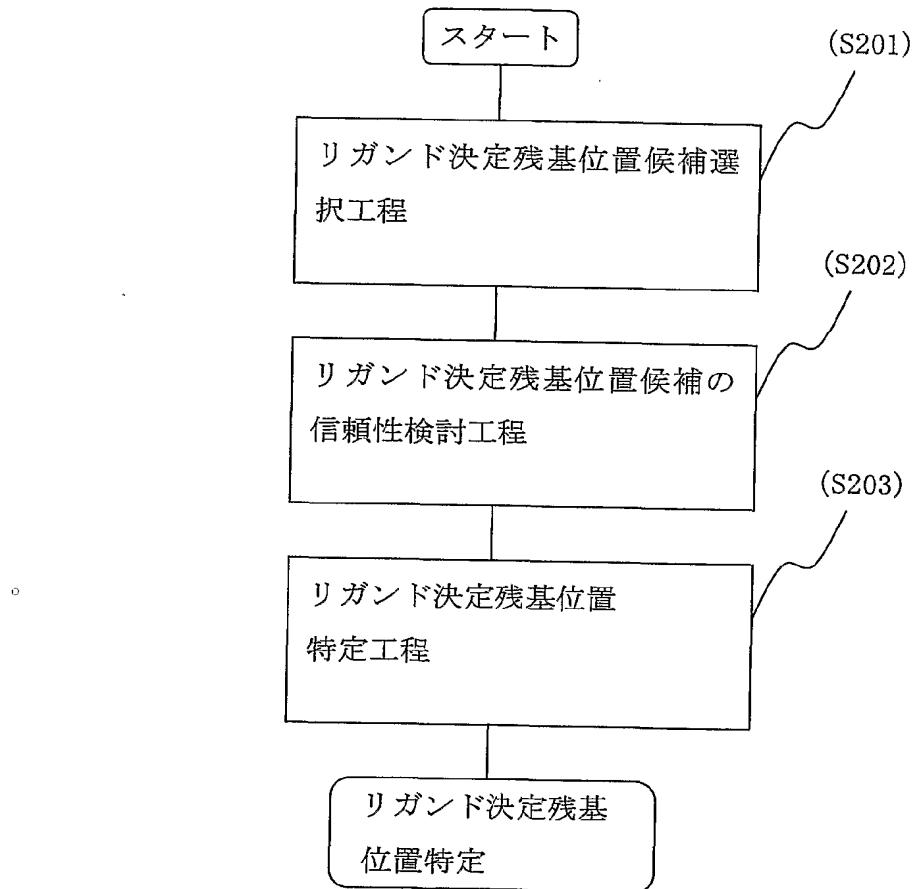
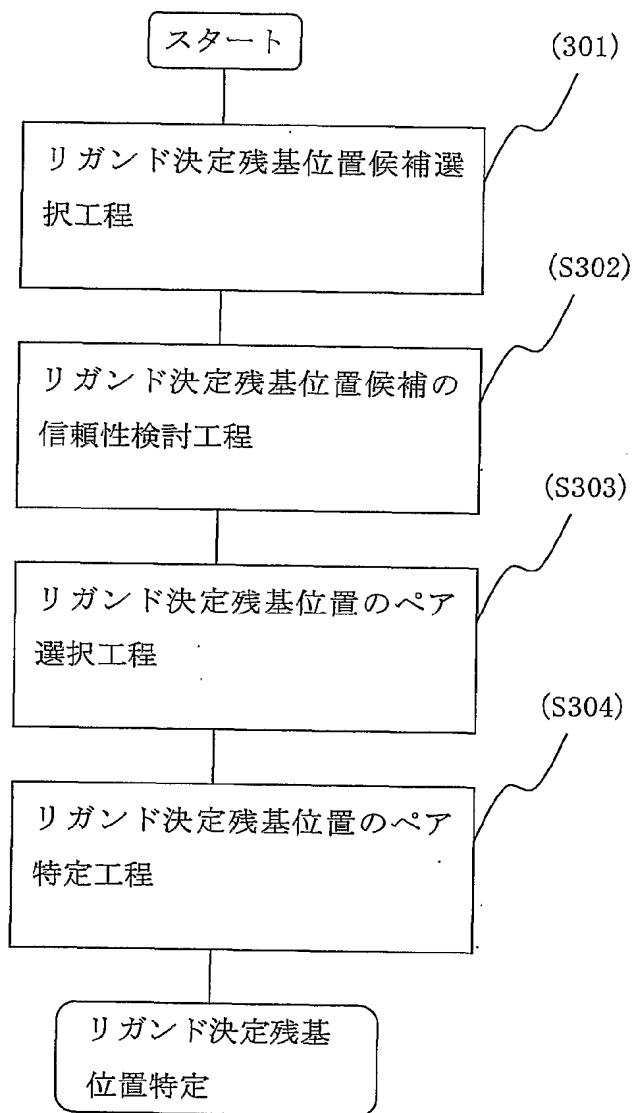


図 3



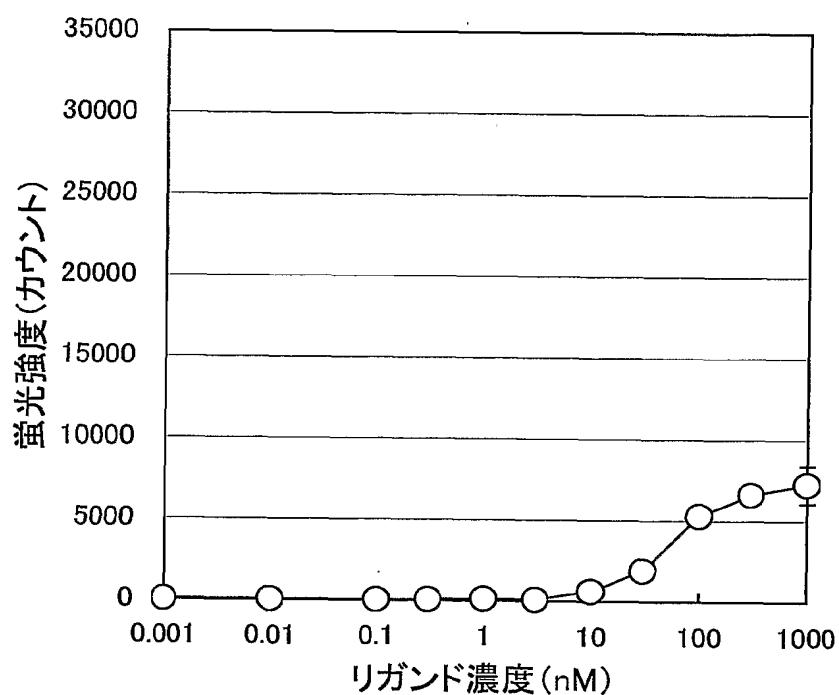
4 / 6

図 4

rhod_bovin	—MNGTEGPN-----FYVPFSNKTGVVRSPEAPQQYLAEPWQFSMLAAYMFLLIM
TGR23-1	MPANFTEGSFDSSGTGQTLDSSPVACTETVTFTEVVEGKEWGSFYYSFKTEQLITLWLF * *** * * ..
rhod_bovin	LGFPINFLTLYVTVQHKKLRPLNYILLNLAVADLFMVFGGFTTLYTSLHGKFVFGPTG
TGR23-1	VFTIVGNSVVLFSTWRRKKKSRTFFVTQLAITDSFTGLVNILTDINWRFTGDFTAPDLV * **..** . * . . * *
rhod_bovin	CNLEGFFATLGGEIALWSLVVLAERYVVVCKPMSNFRGENHAIMGVAFTWWMALACAA
TGR23-1	CRVWRYLQVVLLYASTYVLVLSIDRYHAIVYPMK-FLQGEKQARVLIVIAWSLSFLFSI * **.*.* . * * * * . * . . * ..
rhod_bovin	PPLVGWSRYIPEGMQCSCGIDYYTPHEETNNESFVYMFWHFIPLIVIFFCYGQLVFT
TGR23-1	PTLIIFGKRTLSNGEVQCWALW-PDDSYWTPYMTIVAFLVYF-IPLTISIMYGIVIRT * *. * . . . * . . . * . * * * . * . . * .. *
rhod_bovin	VKEAAAQQQ-----ESATTQAEKEVTRMVIMVI AFLICWL PYAGVAFY
TGR23-1	IWI KSKTYETVISNCSDGKLCSSYNRGLISKA KIKYSIII LAFICCWSPYFLFDIL * * * * *
rhod_bovin	IFTHQGSDFGP FMT-- PAFFAKTSAYNPV Y MMNKQFRNCMVTTLCCGKNPLGDDE
TGR23-1	DNFNLLPDTQERFYASVIIQNLPALNSAINPLIYCVFSSS S-----FPCREQ-RSQDS * . * . * * *
rhod_bovin	ASTTVSKTETSQVAPA-----
TGR23-1	RMTFRERTERHEMQILSKPEFI * . . * . .

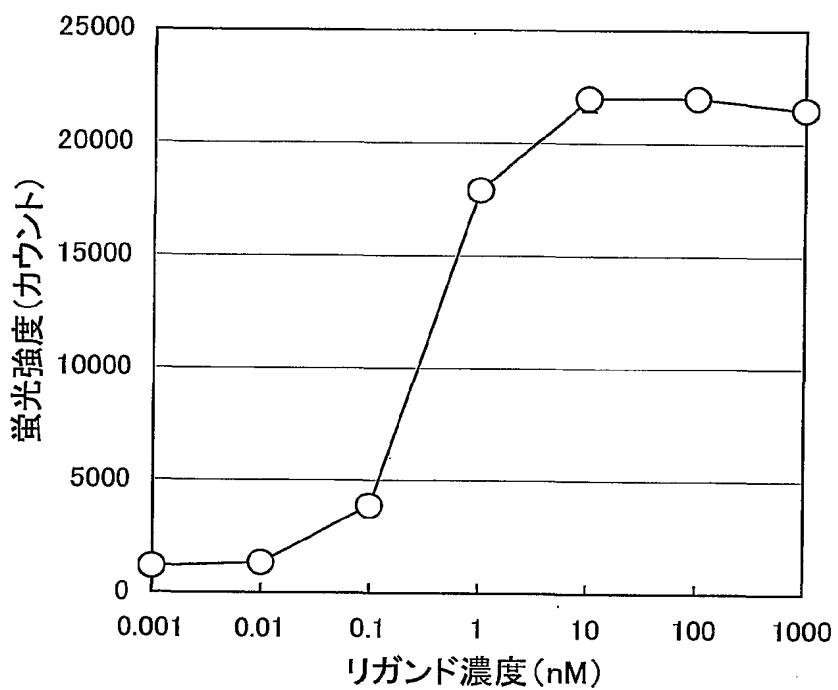
5 / 6

図 5



6 / 6

図 6



1/21
SEQUENCE LISTING

<110> Takeda Chemical Industries, Ltd.

<120> Prediction method of ligand

<130> P02-0089PCT

<160> 40

<210> 1

<211> 18

<212> PRT

<213> Rat

<400> 1

Ser Phe Arg Asn Gly Val Gly Ser Gly Val Lys Lys Thr Ser Phe Arg

5

10

15

Arg Ala

<210> 2

<211> 15

<212> PRT

<213> Rat

<400> 2

Ser Phe Arg Asn Gly Val Gly Ser Gly Val Lys Lys Thr Ser Phe

5

10

15

2/21

<210> 3

<211> 14

<212> PRT

〈213〉 Rat

〈400〉 3

Ser Phe Arg Asn Gly Val Gly Ser Gly Val Lys Lys Thr Ser

5

10

14

210 4

<211> 24

<212> DNA

⟨213⟩ Artificial Sequence

220

〈223〉 Primer

〈400〉 4

cagattttgg gaagtccaaa atga

24

〈210〉 5

<211> 24

<212> DNA

<213> Artificial Sequence

220

〈223〉 Primer

<400> 5

3/21

gagtacgta gtcacactct acag 24

<210> 6

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 6

agatttaattc cccgagtcct ttgc 24

<210> 7

<211> 322

<212> DNA

<213> Human

<400> 7

cagattttgg gaagtccaaa atgatttagct cagtaaaact caatctcatc cttagttctgt	60
cgctgtccac aatgcattgtg ttttgggtt atccaggttcc atcttctaag gtgtctggaa	120
aatctgatta ctttctcatt ctgcataaca gctgcccaac cagattggac aggagcaaag	180
aactagcttt tctaaagcca attttggaga agatgtttgt gaaaagggtcc ttgcataatg	240
gagttggcac agggatgaaa aaaacttcct ttcaaagagc aaaatcatga ctaagtgtgc	300
aaaggactcg ggaaattaat ct	322

<210> 8

<211> 89

<212> PRT

4/21

<213> Human

<400> 8

Met Ile Ser Ser Val Lys Leu Asn Leu Ile Leu Val Leu Ser Leu Ser
5 10 15

Thr Met His Val Phe Trp Cys Tyr Pro Val Pro Ser Ser Lys Val Ser
20 25 30

Gly Lys Ser Asp Tyr Phe Leu Ile Leu Leu Asn Ser Cys Pro Thr Arg
35 40 45

Leu Asp Arg Ser Lys Glu Leu Ala Phe Leu Lys Pro Ile Leu Glu Lys
50 55 60

Met Phe Val Lys Arg Ser Phe Arg Asn Gly Val Gly Thr Gly Met Lys
65 70 75 80

Lys Thr Ser Phe Gln Arg Ala Lys Ser
85

<210> 9

<211> 18

<212> PRT

<213> Human

<400> 9

Ser Phe Arg Asn Gly Val Gly Thr Gly Met Lys Lys Thr Ser Phe Gln
5 10 15

Arg Ala

<210> 10

<211> 15

<212> PRT

5/21

<213> Human

<400> 10

Ser Phe Arg Asn Gly Val Gly Thr Gly Met Lys Lys Thr Ser Phe
5 10 15

<210> 11

<211> 14

<212> PRT

<213> Human

<400> 11

Ser Phe Arg Asn Gly Val Gly Thr Gly Met Lys Lys Thr Ser
5 10

<210> 12

<211> 20

<212> PRT

<213> Human

<400> 12

Ser Phe Arg Asn Gly Val Gly Thr Gly Met Lys Lys Thr Ser Phe Gln
5 10 15
Arg Ala Lys Ser
20

<210> 13

<211> 24

<212> DNA

6/21

<213> Artificial Sequence

<220>

<223> Primer

<400> 13

ccagtcacac aggagggatc tcaa 24

<210> 14

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 14

gcacatcagt cacactctac atag 24

<210> 15

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 15

agattaattc ccagagtctt ttgc 24

7/21

<210> 16

<211> 443

<212> DNA

<213> Mouse

<400> 16

ccagtcacac aggagggatc tcaatgacat tttacttct gaactttct aatataaaag	60
ggccacccaa gcaggctcg acagcaaacg tgaggaaatt ggcaataaaa acccatctgc	120
gcaggtctcg gaaaatccaa aatgattggc tcgttaaaac tcagctcgt cttagctcg	180
tcgctgtctg taatgcacgt gcittggtgt tatccgtcc tctttccaa ggtgcctggg	240
aagcctgatt actttctcat ctigctgagc agctgcccag ccaggctgga ggggagcgcac	300
aggctagctt ttctaaagcc aatttggag aagacatcga tgaaaaggtc ctttcgcaac	360
ggagtccgct cagggcgaa aaaaacttcg ttgcagag caaagcaatg aataagtgtg	420
caaaggactc tggaattaa tct	443

<210> 17

<211> 89

<212> PRT

<213> Mouse

<400> 17

Met Ile Gly Ser Leu Lys Leu Ser Phe Val Leu Ala Leu Ser Leu Ser		
---	--	--

5

10

15

Val Met His Val Leu Trp Cys Tyr Pro Val Leu Ser Ser Lys Val Pro		
---	--	--

20

25

30

Gly Lys Pro Asp Tyr Phe Leu Ile Leu Leu Ser Ser Cys Pro Ala Arg		
---	--	--

35

40

45

Leu Glu Gly Ser Asp Arg Leu Ala Phe Leu Lys Pro Ile Leu Glu Lys		
---	--	--

8/21

50 55 60
Thr Ser Met Lys Arg Ser Phe Arg Asn Gly Val Gly Ser Gly Ala Lys
65 70 75 80
Lys Thr Ser Phe Arg Arg Ala Lys Gln
85

<210> 18

<211> 18

<212> PRT

<213> Mouse

<400> 18

Ser Phe Arg Asn Gly Val Gly Ser Gly Ala Lys Lys Thr Ser Phe Arg
5 10 15
Arg Ala

<210> 19

<211> 15

<212> PRT

<213> Mouse

<400> 19

Ser Phe Arg Asn Gly Val Gly Ser Gly Ala Lys Lys Thr Ser Phe
5 10 15

<210> 20

<211> 14

<212> PRT

<213> Mouse

9/21

<400> 20

Ser Phe Arg Asn Gly Val Gly Ser Gly Ala Lys Lys Thr Ser
5 10

<210> 21

<211> 20
<212> PRT
<213> Mouse

<400> 21

Ser Phe Arg Asn Gly Val Gly Ser Gly Ala Lys Lys Thr Ser Phe Arg
5 10 15
Arg Ala Lys Gln
20

<210> 22

<211> 24
<212> DNA
<213> Artificial Sequence

<220>

<223> Primer

<400> 22

ctgattactt tctcatyytg ctga 24

<210> 23

<211> 199

10/21

<212> DNA

<213> Rat

<400> 23

ctgattactt tctcattttgc tgagtagccct gcccagccagg gctggagggg agcgacggc	60
tagctttctt aaagccaatt ttggagaaga cgtcgatgaa aaggtccctt cgcaacggag	120
tccggctcagg ggtgaaaaaa acttcatttc gaagagcaaa gcaatgaata agtgtgcaaa	180
ggactctggg aattaatct	199

<210> 24

<211> 54

<212> PRT

<213> Rat

<400> 24

Asp Tyr Phe Leu Ile Leu Leu Ser Thr Cys Pro Ala Arg Leu Glu Gly

5 10 15

Ser Asp Gly Leu Ala Phe Leu Lys Pro Ile Leu Glu Lys Thr Ser Met

20 25 30

Lys Arg Ser Phe Arg Asn Gly Val Gly Ser Gly Val Lys Lys Thr Ser

35 40 45

Phe Arg Arg Ala Lys Gln

50

<210> 25

<211> 20

<212> PRT

<213> Mouse

11/21

<400> 25

Ser Phe Arg Asn Gly Val Gly Ser Gly Val Lys Lys Thr Ser Phe Arg

5

10

15

Arg Ala Lys Gln

20

<210> 26

<211> 16

<212> PRT

<213> Human

<400> 26

Ser Phe Arg Asn Gly Val Gly Thr Gly Met Lys Lys Thr Ser Phe Gln

5

10

15

◦

<210> 27

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 27

cttaacaaga acaaaaggcc acag 24

<210> 28

<211> 26

<212> DNA

12/21

<213> Artificial Sequence

<220>

<223> Primer

<400> 28

ttattcattg ctttgctctt cgaaat 26

<210> 29

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 29

ccacccaaggc aggctcagac agcgag 26

<210> 30

<211> 353

<212> DNA

<213> Rat

<400> 30

ccacccaaggc aggctcagac agcgagcgtg aggaatttgg caataaaaac ccatctgcac	60
agatctcgga aaatccaaaa tgattggctc attaaaactc aacctcatct tagctctgtc	120
gctgtccgtg gtacacgtga tttggagttt tccggtcctc tcttccaagg tgcctggaa	180
gcctgattac tttctcattt tgctgagttt ctgccccagcc aggctggagg ggagcgacgg	240

13/21

gctagcttt ctaaagccaa ttttgagaa gacgtcgatg aaaaggcct ttgcAACGG 300
agtccggctca ggggtgaaaa aaacttcatt tcgaagagca aagcaatgaa taa 353

<210> 31

<211> 89

<212> PRT

<213> Rat

<400> 31

Met Ile Gly Ser Leu Lys Leu Asn Leu Ile Leu Ala Leu Ser Leu Ser
5 10 15
Val Val His Val Ile Trp Ser Tyr Pro Val Leu Ser Ser Lys Val Pro
20 25 30
Gly Lys Pro Asp Tyr Phe Leu Ile Leu Leu Ser Thr Cys Pro Ala Arg
35 40 45
Leu Glu Gly Ser Asp Gly Leu Ala Phe Leu Lys Pro Ile Leu Glu Lys
50 55 60
Thr Ser Met Lys Arg Ser Phe Arg Asn Gly Val Gly Ser Gly Val Lys
65 70 75 80
Lys Thr Ser Phe Arg Arg Ala Lys Gln
85

<210> 32

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

14/21

<400> 32

tatagtcgac atgccagcca acttcac 27

<210> 33

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 33

tgtcactagt ctagatgaat tctggctt 28

<210> 34

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 34

ttcactggag acttcacggc a 21

<210> 35

<211> 22

<212> DNA

15/21

<213> Artificial Sequence

<220>

<223> Primer

<400> 35

tagaggcgta gagcagcaca ac 22

<210> 36

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Probe designed for TaqMan PCR, wherein 5' end is labeled by 6-carboxy-fluorescein (Fam) and 3' end is labeled by 6-carboxy-tetramethyl-rhodamine (TAMRA)

<400> 36

acctgggttg ccgagtggtc cgctatTT 28

<210> 37

<211> 371

<212> PRT

<213> Human

<400> 37

Met Pro Ala Asn Phe Thr Glu Gly Ser Phe Asp Ser Ser Gly Thr Gly

16/21

Gln Thr Leu Asp Ser Ser Pro Val Ala Cys Thr Glu Thr Val Thr Phe
20 25 30
Thr Glu Val Val Glu Gly Lys Glu Trp Gly Ser Phe Tyr Tyr Ser Phe
35 40 45
Lys Thr Glu Gln Leu Ile Thr Leu Trp Val Leu Phe Val Phe Thr Ile
50 55 60
Val Gly Asn Ser Val Val Leu Phe Ser Thr Trp Arg Arg Lys Lys
65 70 75 80
Ser Arg Met Thr Phe Phe Val Thr Gln Leu Ala Ile Thr Asp Ser Phe
85 90 95
Thr Gly Leu Val Asn Ile Leu Thr Asp Ile Asn Trp Arg Phe Thr Gly
100 105 110
Asp Phe Thr Ala Pro Asp Leu Val Cys Arg Val Val Arg Tyr Leu Gln
115 120 125
Val Val Leu Leu Tyr Ala Ser Thr Tyr Val Leu Val Ser Leu Ser Ile
130 135 140
Asp Arg Tyr His Ala Ile Val Tyr Pro Met Lys Phe Leu Gln Gly Glu
145 150 155 160
Lys Gln Ala Arg Val Leu Ile Val Ile Ala Trp Ser Leu Ser Phe Leu
165 170 175
Phe Ser Ile Pro Thr Leu Ile Phe Gly Lys Arg Thr Leu Ser Asn
180 185 190
Gly Glu Val Gln Cys Trp Ala Leu Trp Pro Asp Asp Ser Tyr Trp Thr
195 200 205
Pro Tyr Met Thr Ile Val Ala Phe Leu Val Tyr Phe Ile Pro Leu Thr
210 215 220
Ile Ile Ser Ile Met Tyr Gly Ile Val Ile Arg Thr Ile Trp Ile Lys
225 230 235 240
Ser Lys Thr Tyr Glu Thr Val Ile Ser Asn Cys Ser Asp Gly Lys Leu

17/21

245	250	255
Cys Ser Ser Tyr Asn Arg Gly Leu Ile Ser Lys Ala Lys Ile Lys Ala		
260	265	270
Ile Lys Tyr Ser Ile Ile Ile Leu Ala Phe Ile Cys Cys Trp Ser		
275	280	285
Pro Tyr Phe Leu Phe Asp Ile Leu Asp Asn Phe Asn Leu Leu Pro Asp		
290	295	300
Thr Gln Glu Arg Phe Tyr Ala Ser Val Ile Ile Gln Asn Leu Pro Ala		
305	310	315
Leu Asn Ser Ala Ile Asn Pro Leu Ile Tyr Cys Val Phe Ser Ser Ser		
325	330	335
Ile Ser Phe Pro Cys Arg Glu Gln Arg Ser Gln Asp Ser Arg Met Thr		
340	345	350
Phe Arg Glu Arg Thr Glu Arg His Glu Met Gln Ile Leu Ser Lys Pro		
355	360	365
Glu Phe Ile		
370		

<210> 38

<211> 1113

<212> DNA

<213> Human

<400> 38

atgccagcca acttcacaga gggcagcttc gattccagtg ggaccggca gacgctggat	60
tcttccccag tggcttgcac tgaaacagtg actttactg aagtggtgga aggaaaggaa	120
tgggttcct tctactactc cttaagact gagcaattga taactctgtg ggtcctcttt	180
gttittacca ttgttgaaa ctccgttgtc ctttttcca catggaggag aaagaagaag	240
tcaagaatga ccttcttgt gactcagctg gccatcacag attcttcac aggactggtc	300

18/21

aacatcttga cagatattaa ttggcgattc actggagact tcacggcacc tgacctggtt	360
tgcgcgagtgg tccgctatTT gcaggTTTGTG ctgtctacg cctctaccta cgtccTggTG	420
tccctcagca tagacagata ccatgccatc gtctacccc tgaagtccT tcaaggagaa	480
aagcaagcca gggcctcat tgtgatGCC tggagcctgt ctTTTCTGTT ctccattccc	540
accctgatca tatttggaa gaggacactg tccaacggtg aagtgcagtG ctggccctg	600
tggcctgacg actcctactg gaccccatac atgaccatcg tggccttcT ggtgtactTC	660
atccctctga caatcatcag catcatgtat ggcattgtGA tccgaactat ttggattaaa	720
agcaaaacct acgaaacagt gatttccaac tgctcagatG ggaaactgtG cagcagctat	780
aaccgaggac tcatctaaa ggcaaaaatc aaggctatca agtatacgat catcatcatt	840
tttgccttca tctgctgttg gagtccatac ttccTgtttG acatTTGGA caatttcaac	900
ctccttccag acacccagga gcgtttctat gcctctgtGA tcattcagaa cctgccagca	960
ttgaatagtG ccatcaaccc cctcatctac tttgtcttca gcagctccat ctctttcccc	1020
tgcagggagc aaagatcaca ggattccaga atgacgttcc gggagagaac tgagaggcat	1080
gagatgcaga ttctgtccaa gccagaattc atc	1113

<210> 39

<211> 371

<212> PRT

<213> Human

<400> 39

Met Pro Ala Asn Phe Thr Glu Gly Ser Phe Asp Ser Ser Gly Thr Gly

5 10 15

Gln Thr Leu Asp Ser Ser Pro Val Ala Cys Thr Glu Thr Val Thr Phe

20 25 30

Thr Glu Val Val Glu Gly Lys Glu Trp Gly Ser Phe Tyr Tyr Ser Phe

35 40 45

Lys Thr Glu Gln Leu Ile Thr Leu Trp Val Leu Phe Val Phe Thr Ile

50 55 60

19/21

Val Gly Asn Ser Val Val Leu Phe Ser Thr Trp Arg Arg Lys Lys Lys
 65 70 75 80
 Ser Arg Met Thr Phe Phe Val Thr Gln Leu Ala Ile Thr Asp Ser Phe
 85 90 95
 Thr Gly Leu Val Asn Ile Leu Thr Asp Ile Ile Trp Arg Phe Thr Gly
 100 105 110
 Asp Phe Thr Ala Pro Asp Leu Val Cys Arg Val Val Arg Tyr Leu Gln
 115 120 125
 Val Val Leu Leu Tyr Ala Ser Thr Tyr Val Leu Val Ser Leu Ser Ile
 130 135 140
 Asp Arg Tyr His Ala Ile Val Tyr Pro Met Lys Phe Leu Gln Gly Glu
 145 150 155 160
 Lys Gln Ala Arg Val Leu Ile Val Ile Ala Trp Ser Leu Ser Phe Leu
 165 170 175
 Phe Ser Ile Pro Thr Leu Ile Ile Phe Gly Lys Arg Thr Leu Ser Asn
 180 185 190
 Gly Glu Val Gln Cys Trp Ala Leu Trp Pro Asp Asp Ser Tyr Trp Thr
 195 200 205
 Pro Tyr Met Thr Ile Val Ala Phe Leu Val Tyr Phe Ile Pro Leu Thr
 210 215 220
 Ile Ile Ser Ile Met Tyr Gly Ile Val Ile Arg Thr Ile Trp Ile Lys
 225 230 235 240
 Ser Lys Thr Tyr Glu Thr Val Ile Ser Asn Cys Ser Asp Gly Lys Leu
 245 250 255
 Cys Ser Ser Tyr Asn Arg Gly Leu Ile Ser Lys Ala Lys Ile Lys Ala
 260 265 270
 Ile Lys Tyr Ser Ile Ile Ile Leu Ala Phe Ile Cys Cys Trp Ser
 275 280 285
 Pro Tyr Phe Leu Phe Asp Ile Leu Asp Asn Phe Asn Leu Leu Pro Asp

20/21

290	295	300
Thr Gln Glu Arg Phe Tyr Ala Ser Val Ile Ile Gln Asn Leu Pro Ala		
305	310	315
Leu Asn Ser Ala Ile Asn Pro Leu Ile Tyr Cys Val Phe Ser Ser Ser		
	325	330
		335
Ile Ser Phe Pro Cys Arg Glu Arg Arg Ser Gln Asp Ser Arg Met Thr		
	340	345
Phe Arg Glu Arg Thr Glu Arg His Glu Met Gln Ile Leu Ser Lys Pro		
	355	360
Glu Phe Ile		365
370		

<210> 40

<211> 1113

<212> DNA

<213> Human

<400> 40

atgccagcca acttcacaga gggcagcttc gatccagtg ggaccggca gacgctggat	60
tcttccccag tggcttgcac tgaaacagtg actttactg aagtggtgga aggaaaggaa	120
tgggttctct tctactactc cttaagact gagcaattga taactctgtg ggtcctctt	180
gtttttacca ttgttggaaa ctccgttgtc ctttttcca catggaggag aaagaagaag	240
tcaagaatga ccttccttgt gactcagctg gccatcacag attcttcac aggactggc	300
aacatcttga cagatattat ttggcgattc actggagact tcacggcacc tgacctgg	360
tgccgagtgg tccgctatcc gcaggttgtc ctgctctacg cctctaccta cgtcctgg	420
tccctcagca tagacagata ccatgccatc gtctacccca tgaagtccct tcaaggagaa	480
aagcaagcca gggcctcat tgtgatgcc tggagcctgt ctttctgtt ctccattccc	540
accctgatca tatttggaa gaggacactg tccaacggtg aagtgcagtg ctggccctg	600
tggcctgacg actcctactg gacccatac atgaccatcg tggccttct ggtgtacttc	660

21/21

atccctctga caatcatcag catcatgtat ggcattgtga tccgaactat ttggattaaa	720
agcaaaaacct acgaaacagt gatitccaac tgctcagatg ggaaactgtg cagcagctat	780
aaccgaggac tcatctcaa ggcaaaaatc aaggctatca agtatacgat catcatcatt	840
cttgccttca tctgcgtttg gagtcatac ttccgtttt acattttggaa caatttcaac	900
cgccttcag acacccagga gcgttttat gcctctgtga tcattcagaa cctgccagca	960
ttgaatagtgc ccatcaaccc cctcatctac tgggtttca gcagctccat ctctttcccc	1020
tgcagggagc gaagatcaca ggattccaga atgacgttcc gggagagaac cgagaggcat	1080
gagatgcaga ttctgtccaa gccagaattc atc	1113